

# SEARCH REQUEST FORM

Requestor's Name: \_\_\_\_\_ Serial Number: \_\_\_\_\_  
Date: \_\_\_\_\_ Phone: \_\_\_\_\_ Art Unit: \_\_\_\_\_

## Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

## STAFF USE ONLY

Date completed: 01-10-03  
Searcher: Beverly E. 4494  
Terminal time: 20  
Elapsed time: \_\_\_\_\_  
CPU time: \_\_\_\_\_  
Total time: 23  
Number of Searches: \_\_\_\_\_  
Number of Databases: 1

### Search Site

\_\_\_\_ STIC  
\_\_\_\_ CM-1  
\_\_\_\_ Pre-S

### Type of Search

\_\_\_\_ N.A. Sequence  
\_\_\_\_ A.A. Sequence  
\_\_\_\_ Structure  
\_\_\_\_ Bibliographic

### Vendors

\_\_\_\_ IG  
\_\_\_\_ STN  
\_\_\_\_ Dialog  
\_\_\_\_ APS  
\_\_\_\_ Geninfo  
\_\_\_\_ SDC  
\_\_\_\_ DARC/Questel  
☒ Other CGN



Fri Jan 10 09:31:31 2003

us-09-873-737a-5.rge

Page 1

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 18:53:20 (Search time 5918 Seconds  
17074.171 Million cell updates/sec)

Title: US-09-873-737a-5  
Perfect score: 3472  
Sequence: 1 gggcgattggcgccgacgc.....taaacataatgagtcgcc 3472

Scoring table: IDENTITY-NUC  
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: GenEmbl:  
1: gb\_ha:\*  
2: gb\_hg:\*  
3: gb\_in:\*  
4: gb\_ov:\*  
5: gb\_ov:\*  
6: gb\_ov:\*  
7: gb\_ov:\*  
8: gb\_ov:\*  
9: gb\_ov:\*  
10: gb\_ov:\*  
11: gb\_ov:\*  
12: gb\_ov:\*  
13: gb\_ov:\*  
14: gb\_ov:\*  
15: gb\_ov:\*  
16: gb\_ov:\*  
17: gb\_ov:\*  
18: gb\_ov:\*  
19: gb\_ov:\*  
20: gb\_ov:\*  
21: gb\_ov:\*  
22: gb\_ov:\*  
23: gb\_ov:\*  
24: gb\_ov:\*  
25: gb\_ov:\*  
26: gb\_ov:\*  
27: gb\_ov:\*  
28: gb\_ov:\*  
29: gb\_ov:\*  
30: gb\_ov:\*  
31: gb\_ov:\*  
32: gb\_ov:\*  
33: gb\_ov:\*  
34: gb\_ov:\*  
35: gb\_ov:\*  
36: gb\_ov:\*  
37: gb\_ov:\*  
38: gb\_ov:\*  
39: gb\_ov:\*  
40: gb\_ov:\*  
41: gb\_ov:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3359	96.7	3362	9 AF104260	AF104260 Homo sapi
2	3344.6	96.3	3399	9 BC028581	BC028581 Homo sapi
3	3337.6	96.1	3421	9 AF387507	AF387507 Homo sapi
4	2513.6	72.4	2579	9 AK093133	AK093133 Homo sapi
5	2311	66.6	2328	9 AF264004	AF264004 Homo sapi
6	1839.2	53.0	3910	10 AB032604	AB032604 Mus muscu
7	1839.2	53.0	4064	10 AF384055	AF384055 Mus muscu
8	1067.2	23.9	3273	5 AF336369	AF336369 Homo sapi
9	830.8	23.9	193697	2 AC127071	AC127071 Homo sapi
10	830.8	23.9	198104	2 AC127071	AC127071 Homo sapi
11	679.2	19.6	3006	9 BC031060	BC031060 Homo sapi
12	679.2	19.6	3138	9 AK093475	AK093475 Homo sapi
13	613.2	17.7	2456	9 AK096837	AK096837 Homo sapi
14	565	16.3	3057	9 AY014899	AY014899 Strongylo
15	550.4	15.9	590	6 AX381696	AX381696 Sequence
16	495.8	14.3	2605	6 AY014900	AY014900 Strongylo
17	459.4	13.2	460	6 AX395819	AX395819 Sequence
18	458.4	13.2	4913	10 AB032605	AB032605 Mus muscu
19	456.6	12.6	3610	9 BC025995	BC025995 Homo sapi
20	438.8	12.5	3620	9 AK056418	AK056418 Homo sapi
21	438.2	10.9	1723	3 AY014901	AY014901 Strongylo
22	378.4	10.9	3649	6 AX247684	AX247684 Sequence
23	377	10.9	3659	10 AF285586	AF285586 Mus muscu
24	377	10.9	2272	9 AK027497	AK027497 Homo sapi
25	329	9.5	2761	3 AF145680	AF145680 Drosophila
26	310.2	8.9	2761	3 AF143554	AF143554 Homo sapi
27	253.4	7.3	1733	9 AK001213	AK001213 Homo sapi
28	238.8	6.9	61995	2 AC090565	AC090565 Homo sapi
29	203.4	5.4	7781	2 DMDNRP2	DMDNRP2 Homo sapi
30	186.2	5.3	37980	2 AC005116	AC005116 Drosophila
31	184.6	5.3	78507	2 AC019704	AC019704 Drosophila
32	184.6	5.3	169210	3 AC092226	AC092226 Drosophila
33	184.6	5.3	176306	3 AC007186	AC007186 Drosophila
34	184.6	5.3	276576	3 AE003630	AE003630 Drosophila
35	184.6	5.3	276576	3 AE003630	AE003630 Drosophila
36	183	5.3	901	3 AF334408	AF334408 Sequence
37	170.8	4.9	548	6 AX387649	AX387649 Sequence
38	170.6	4.9	236516	2 AC111089	AC111089 Mus muscu
39	160.4	4.6	3677	2 AC017885	AC017885 Drosophila
40	160.4	4.6	3855	3 AE003107	AE003107 Drosophila
41	141.2	4.1	9359	3 AF104355	AF104355 Drosophila
42	138.4	4.0	39365	2 AC100141	AC100141 Mus muscu
43	129.6	3.7	66611	2 AC090147	AC090147 Homo sapi
44	126.6	3.6	66611	2 AC090147	AC090147 Homo sapi
45	126.4	3.6	1263	9 AK000397	AK000397 Homo sapi

## ALIGNMENTS

RESULT 1  
AF104260 3362 bp mRNA linear PRI 09-JAN-2002  
LOCUS Homo sapiens HIWI mRNA, complete cds.  
DEFINITION AF104260  
ACCESSION AF104260  
VERSION AF104260.2 GI:18098557  
KEYWORDS  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 3362)  
Cox.D.N., Chao.A., Baker,J., Chang,L., Qiao.D. and Lin.H.  
A novel class of evolutionarily conserved genes defined by piwi are  
essential for stem cell self-renewal

Pred. No. is the number of results predicted by chance to have a

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Genes Dev. 12 (23), 3715-3727 (1998)  
99069219  
9851978  
2 (bases 1 to 3362)  
Qiao, D., Zeeman, A.-M., Deng, W., Looijenga, L.H.J. and Lin, H.  
Molecular characterization of hwi1, a human member of the piwi stem  
cell gene family whose overexpression is correlated to seminomas  
Unpublished  
3 (bases 1 to 3362)  
Lin, H.  
Direct Submission  
Submitted (04-NOV-1998) Cell Biology, Duke University Medical  
Center, 412 Nanaline Duke Bldg., Research Dr., Durham, NC 27710,  
USA  
4 (bases 1 to 3362)  
Qiao, D., Zeeman, A.-M., Deng, W., Looijenga, L.H.J. and Lin, H.  
Direct Submission  
Submitted (09-JAN-2002) Cell Biology, Duke University, 412 Nanaline  
Duke Building, DUMC, Durham, NC 27710, USA  
Nucleotide sequence updated by submitter  
On Jan 9, 2002 this sequence version replaced gi:4038412.  
Location/Qualifiers  
1. .3362  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/map="12q24.33"  
/tissue\_type="testis"  
57..2642  
/note="member of the PIWI family"  
/codon\_start=1  
/product="H1WI"  
/protein\_id="AAC97371.2"  
/db\_xref="GI:18098558"  
/translations="MTGRARARARGOETAQIVGSTASQOQVYIQRPQPPAEG  
ELFGRGRQRTAGTAKSGLQISAGFOELSLAERGGRRRDFDLGVNTRQNDHVK  
HAFDGLILPKRLQKQVTEVFSKTRNGEDVRIITLTNELPSTPCLOFVNIIFR  
LLKMLQIGRNYNPNDPIDIPSHRLVWPGFTSIIQYENSIMLCDDVSHKVLRS  
ETVLDPMFNFYHOTSEHKFQEQSVKELIGLVLTNINIKINIKITLPPVDDIDQNPSTFKL  
TDMRDNFNKDLAVHRLPPEORQEVKGLIDYIHKNDVORELRDWSGLSFDPSNLL  
SFSGRIQLTKHOQKTFDYNFQFADSKETRGAPLISVKPLDNLWLIYTRRYEAA  
NSLIQNLKVPAMPQMKRKAIMEVDDETEAYLRVLOKVTADTOIVVCLLSNRKD  
KYDAIKYLCDCPTSCQVVARTLGKQVTAIAIKIALOMNCKMKGELWRVDPILK  
LVMIIGDICHDMTAGRSIAGFVASINEGTRWFSRCIFODRGOELVDGLKVLQAA  
LRWNSCNEYMPSRIIYRVDGVDGQKTLVNVVEPFLDCLKSIKRGVPRLTIVV  
KRWNTFRFAAGSRLQNPFGTVIDVEVTRPEWDFEIVSQVRSVSPHTYIVY  
DNSGLKPDHIOQLTKLCHYIYNWPGVIRVPAPQYAKLAFVQGSIHREPNSLSN  
RLYLI"

source

CDS

BASE COUNT 1026 a 659 c 759 g 918 t  
ORIGIN  
Query Match  
Best Local Similarity 96.78; Score 3359; DB 9; Length 3362;  
Matches 3359; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 111 CTGAGGTCACAGGACAGGACTAGGCGGAGGCGGCTCAAGAAATAGAAACAATGA 170  
DB 1 CTGAGGTCACAGGACAGGACTAGGCGGAGGCGGCTCAAGAAATAGAAACAATGA 60  
QY 171 CTGGAGAGCCGAGCAGAGCCAGAGGAGGCGGCGGCTCAGGACAGGCGAGCTGG 230  
DB 61 CTGGAGAGCCGAGCAGAGCCAGAGGAGGCGGCGGCTCAGGACAGGCGAGCTGG 120  
QY 231 TGGGCTCCACTGCGCTAGCAACTGTTATATTCAGCCTAGGCTCAGCGCCACGAG 290  
DB 121 TGGGCTCCACTGCGCTAGCAACTGTTATATTCAGCCTAGGCTCAGCGCCACGAG 180  
QY 291 CAGAGGGGGAATTTATTTGGCCGTGAGCGGACAGAGAGGACAGCAGAGGAAACAGCAAGT 350  
DB 181 CAGAGGGGGAATTTATTTGGCCGTGAGCGGACAGAGAGGAAACAGCAGGAAACAGCAAGT 240

QY 351 CACAAGGACTCCAGATATCTGCTGATTTTCAGAGTTATTCAGAGGAGGAGGATC 410  
DB 241 CACAAGGACTCCAGATATCTGCTGATTTTCAGAGTTATTCAGAGGAGGAGGATC 300  
QY 411 GTCGCTAGAGATTTTCATGATCTTTGGTGTGATACATACAGGAGGAGGAGGATC 470  
DB 301 GTCGCTAGAGATTTTCATGATCTTTGGTGTGATACATACAGGAGGAGGAGGATC 360  
QY 471 AATCAAAAACAGGTTCTTCAGGCAATTTATAGTTAAGGTTAAGCACTAACCATTCGGGCTGA 530  
DB 361 AATCAAAAACAGGTTCTTCAGGCAATTTATAGTTAAGGTTAAGCACTAACCATTCGGGCTGA 420  
QY 531 CATCCGCTCCAGTGGGCTTTATATCATGATCATCATGATGACTATATACCCCACTGATGAAG 590  
DB 421 CATCCGCTCCAGTGGGCTTTATATCATGATCATCATGATGACTATATACCCCACTGATGAAG 480  
QY 591 CGAAGAAGCTCCGTTTCAGTCTCTTTTTCACACGAGGATCTAATTTGAAAAGTGTGATG 650  
DB 481 CGAAGAAGCTCCGTTTCAGTCTCTTTTTCACACGAGGATCTAATTTGAAAAGTGTGATG 540  
QY 651 CTTTGTATGGAACGATATTTTACCTAAAAGACTACAGCAAAAGGTTTACTGAAGTTT 710  
DB 541 CTTTGTATGGAACGATATTTTACCTAAAAGACTACAGCAAAAGGTTTACTGAAGTTT 600  
QY 711 TTAGTAAGACCCGGAATGGAGAGGATGTGAGGATTAACGATCACTTTAAACAATGAATTC 770  
DB 601 TTAGTAAGACCCGGAATGGAGAGGATGTGAGGATTAACGATCACTTTAAACAATGAATTC 660  
QY 771 CACCTACATCAACCACTCTGTTTGCAGTTCTTATAATTTTTCAGGAGGCTTTGAAA 830  
DB 661 CACCTACATCAACCACTCTGTTTGCAGTTCTTATAATTTTTCAGGAGGCTTTGAAA 720  
QY 831 TCATGATTTTGCACAAATTTGACGAAATTTATATAACCAATGACCCAAATGATATTC 890  
DB 721 TCATGATTTTGCACAAATTTGACGAAATTTATATAACCAATGACCCAAATGATATTC 780  
QY 891 CAAGTCAAGGTTGTGATTTGGCTGGCTCTACTTTCATCTCTTCAAGTGAAGTCTTTGGATT 950  
DB 781 CAAGTCAAGGTTGTGATTTGGCTGGCTCTACTTTCATCTCTTCAAGTGAAGTCTTTGGATT 840  
QY 951 GCATCATCTCTGACACTGACGTTAGGCAATTAAGTCTTCAAGTGAAGTCTTTGGATT 1010  
DB 841 GCATCATCTCTGACACTGACGTTAGGCAATTAAGTCTTCAAGTGAAGTCTTTGGATT 900  
QY 1011 TCATGTTCAACTTTTATCATCACAGAGAGACATAAATTTCAAGAAACAAGTTTCCAAG 1070  
DB 901 TCATGTTCAACTTTTATCATCACAGAGAGACATAAATTTCAAGAAACAAGTTTCCAAG 960  
QY 1071 AANTAATAGGTTTGTGTTTCTTACCAAGTATAACATTAAGACATACAGAGTGGATGATA 1130  
DB 961 AANTAATAGGTTTGTGTTTCTTACCAAGTATAACATTAAGACATACAGAGTGGATGATA 1020  
QY 1131 TTGACTGGGACCAAGATCCCAAGAGCCTTTTAAGAAAGCGGCTCTGGGGTCAAGT 1190  
DB 1021 TTGACTGGGACCAAGATCCCAAGAGCCTTTTAAGAAAGCGGCTCTGGGGTCAAGT 1080  
QY 1191 TCTTGAATATCTACAGGAGCAATACCAAGAGATCAACGAGTCAAGGAGGCTCTGCT 1250  
DB 1081 TCTTGAATATCTACAGGAGCAATACCAAGAGATCAACGAGTCAAGGAGGCTCTGCT 1140  
QY 1251 TGGTCAAGCCGACCAAGAGAGGCGGGGCTTGGGGGACACTGCCAGGCGCTGCCATGC 1310  
DB 1141 TGGTCAAGCCGACCAAGAGAGGCGGGGCTTGGGGGACACTGCCAGGCGCTGCCATGC 1200  
QY 1311 TCATTCCTGAGCTCTGCTATCTTACAGGCTTAACTGATAAATCGTAATGATTTAAACG 1370  
DB 1201 TCATTCCTGAGCTCTGCTATCTTACAGGCTTAACTGATAAATCGTAATGATTTAAACG 1260  
QY 1371 TGATGAAAGACTTAGCGCTTCTACAAAGCTAACTCCAGAGCAAGGAGGAGGAGTGG 1430  
DB 1261 TGATGAAAGACTTAGCGCTTCTACAAAGCTAACTCCAGAGCAAGGAGGAGGAGTGG 1320  
QY 1431 GACGACTCATGATTACATTTCAAAAAACGATATGTTCAAGGGAGCTTCGAGAGCTGGG 1490

Db 2401 AGGCTGTGAGAAAGTGTAGTCTTCTCCACACATTAACATGTATCTTANGACACAGCG 2460

Qy 2571 GCCTGAGCAGACACACATACAGCGCTTGAACCTACAGCTGTGCACATCTATTACAACT 2630

Db 2461 GCCTGAGCAGACACACATACAGCGCTTGAACCTACAGCTGTGCACATCTATTACAACT 2520

Qy 2631 GGCAGAGTGCATTCGTGTCTCTGTCTCTGCAAGTGGCCCAAGCTGTCTTCTTG 2690

Db 2521 GGCAGAGTGCATTCGTGTCTCTGTCTCTGCAAGTGGCCCAAGCTGTCTTCTTG 2580

Qy 2691 TTGGCAGAGTATTCACAGAGGCAATCTGTCACTGTCAACCGCCTTTATCTACCT 2750

Db 2581 TTGGCAGAGTATTCACAGAGGCAATCTGTCACTGTCAACCGCCTTTATCTACCT 2640

Qy 2751 AACCTGCAGAAAGCAGTACAGCGCTTCTCTTTGAATGACCTTTGGATTTTTTANG 2810

Db 2641 AACCTGCAGAAAGCAGTACAGCGCTTCTCTTTGAATGACCTTTGGATTTTTTANG 2700

Qy 2811 CTTTATTTACTTTTTTTTTTAACCTGTATCTTCTGTGATGAACCTTGGAGAGGGATTTAG 2870

Db 2701 CTTTATTTACTTTTTTTTTTAACCTGTATCTTCTGTGATGAACCTTGGAGAGGGATTTAG 2760

Qy 2871 GAGATCTAGCATTTTATTTCTAGCATTCGTATTCACCGGCTTCTTATTTATATGTA 2930

Db 2761 GAGATCTAGCATTTTATTTCTAGCATTCGTATTCACCGGCTTCTTATTTATATGTA 2820

Qy 2931 AATTAGATTTTATTTTATTTTATCTTCTGTCTCTGATAGATTTTGTGAGCATTTTTTGTG 2990

Db 2821 AATTAGATTTTATTTTATTTTATCTTCTGTCTCTGATAGATTTTGTGAGCATTTTTTGTG 2880

Qy 2991 TTTATTTTGAAGAAATGTGATAGATCTGTGTAGTATTAACAGACTCTCGAGAGTA 3050

Db 2881 TTTATTTTGAAGAAATGTGATAGATCTGTGTAGTATTAACAGACTCTCGAGAGTA 2940

Qy 3051 TTTGAATGTGTGTGAGATTTTACTTAAGCTACTTTCAGAGTGAAGAGTCTTACTTA 3110

Db 2941 TTTGAATGTGTGTGAGATTTTACTTAAGCTACTTTCAGAGTGAAGAGTCTTACTTA 3000

Qy 3111 TTAACCTATTTTACTTTATTTTTGAGATACCTGTGTGAATTTTGAATTTTGAAGATTAAGGC 3170

Db 3001 TTAACCTATTTTACTTTATTTTTGAGATACCTGTGTGAATTTTGAAGATTAAGGC 3060

Qy 3171 GTAAAGTGAAGTGTCTACACAGATTAAGAGTGAAGATTAAGTCTTAAAGATTA 3230

Db 3061 GTAAAGTGAAGTGTCTACACAGATTAAGAGTGAAGATTAAGTCTTAAAGATTA 3120

Qy 3231 AGTACTTATTTATTAACCTATTCACAGATTAAGAGTGAAGATTAAGTCTTAAAGATTA 3290

Db 3121 AGTACTTATTTATTAACCTATTCACAGATTAAGAGTGAAGATTAAGTCTTAAAGATTA 3180

Qy 3291 GCAGGGGGCTAGTGTGTATGGAGTAAAAAAACATTTGAAATTTTAAATGTCCAA 3350

Db 3181 GCAGGGGGCTAGTGTGTATGGAGTAAAAAAACATTTGAAATTTTAAATGTCCAA 3240

Qy 3351 GAAACATTTTAAACCTTTAACAAAAAGCCATGATTAATCTCTATTTTAAATCAATCAC 3410

Db 3241 GAAACATTTTAAACCTTTAACAAAAAGCCATGATTAATCTCTATTTTAAATCAATCAC 3300

Qy 3411 TATTATTTTGTGTGAACTGGAGATTAATCTTATTTGTATTAATAATTAATGAGTGC 3470

Db 3301 TATTATTTTGTGTGAACTGGAGATTAATCTTATTTGTATTAATAATTAATGAGTGC 3360

Qy 3471 CC 3472

Db 3361 CC 3362

RESULT 2 3399 bp mRNA linear PRI 25-APR-2002

BC028581 BC028581 LOCUS Homo sapiens, plwi-like 1 (Drosophila), clone MGC:26748

DEFINITION IMAGE:4827445, mRNA, complete cds.

ACCESSION BC028581

RESULT 2			
BC028581	3399 bp	mrna	PRI 25-APR-2002
LOCUS	BC028581		
DEFINITION	Homo sapiens, piwi-like 1 (Drosophila), clone WGC:26748		
ACCESSION	IMAGE:4827445, mRNA, complete cds.		
	BC028581		









QY 1304 GCGATGCTCATTCTGAGCTCTGCTATCTTACAGCTCTTAACTGATTAATGCGTATGAT 1363  
 Db 1201 GCGATGCTCATTCTGAGCTCTGCTATCTTACAGCTCTTAACTGATTAATGCGTATGAT 1260  
 QY 1364 TTTAAGCTGATGAAAGACTTACCGCTTATACAGAGCTTAACTGAGGAAAGGAGCGCT 1423  
 Db 1261 TTTAAGCTGATGAAAGACTTACCGCTTATACAGAGCTTAACTGAGGAAAGGAGCGCT 1320  
 QY 1424 GAAATGGAGAGCTCATTGATTTACATTAATTAATTAATTAATTAATTAATTAATTAAT 1483  
 Db 1321 GAAATGGAGAGCTCATTGATTTACATTAATTAATTAATTAATTAATTAATTAATTAAT 1380  
 QY 1484 GACTGGGCTTGAAGCTTGAATTCCTTACCTGCTTCTGAGAGAAATTTTGCAACA 1543  
 Db 1381 GACTGGGCTTGAAGCTTGAATTCCTTACCTGCTTCTGAGAGAAATTTTGCAACA 1440  
 QY 1544 GAAAGATTCACCAAGGTGAGAAACATTTGATTTACATTCATCAATTTGAGATGCTG 1603  
 Db 1441 GAAAGATTCACCAAGGTGAGAAACATTTGATTTACATTCATCAATTTGAGATGCTG 1500  
 QY 1604 AAGGAAACAAAGAGTGCACATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1663  
 Db 1501 AAGGAAACAAAGAGTGCACATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1560  
 QY 1664 TATACGGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1723  
 Db 1561 TATACGGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1620  
 QY 1724 CAGGCAATGGGCAATGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1783  
 Db 1621 CAGGCAATGGGCAATGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1680  
 QY 1784 GCTTACTTAAAGATCTTAACTGAGCAAAAGTTCACAGACACCCAGATTAATTAATTAAT 1843  
 Db 1681 GCTTACTTAAAGATCTTAACTGAGCAAAAGTTCACAGACACCCAGATTAATTAATTAAT 1740  
 QY 1844 TTGTCAAGTAATTCGAGAGCAAAATACGATGCTTAAATTAATTAATTAATTAATTAATTAAT 1903  
 Db 1741 TTGTCAAGTAATTCGAGAGCAAAATACGATGCTTAAATTAATTAATTAATTAATTAATTAAT 1800  
 QY 1904 CCTACCCCAAGTCAAGT 1963  
 Db 1801 CCTACCCCAAGTCAAGT 1860  
 QY 1964 ATTGCTAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2023  
 Db 1861 ATTGCTAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1920  
 QY 2024 GACATCCCTGAGAGCTGATGATTCGTTGGCATCGATGTTTACATGACATGACATGACAT 2083  
 Db 1921 GACATCCCTGAGAGCTGATGATTCGTTGGCATCGATGTTTACATGACATGACATGACAT 1980  
 QY 2084 GGGGCGAGGTCAATGCGAGATTTGTGCGAGCATCAATGAAGGATGACCGCTGTGCTT 2143  
 Db 1981 GGGGCGAGGTCAATGCGAGATTTGTGCGAGCATCAATGAAGGATGACCGCTGTGCTT 2040  
 QY 2144 TCAGCTGATTAATTCAGAGATGAGAGACAGAGCTGTGATGATGAGCTTAAAGTCTGCTG 2203  
 Db 2041 TCAGCTGATTAATTCAGAGATGAGAGACAGAGCTGTGATGATGAGCTTAAAGTCTGCTG 2100  
 QY 2204 CAAAGGCTCTGAGAGCTTGGATTAATGCTCAATGATGATGATGATGATGATGATGATGAT 2263  
 Db 2101 CAAAGGCTCTGAGAGCTTGGATTAATGCTCAATGATGATGATGATGATGATGATGATGAT 2160  
 QY 2264 TACCGGCAATGCGATGAGAGAGCGGCACTGTAACCACTGTGTGAATCTACGAGAGTCCACAG 2323  
 Db 2161 TACCGGCAATGCGATGAGAGAGCGGCACTGTAACCACTGTGTGAATCTACGAGAGTCCACAG 2220  
 QY 2324 TTTTGGATTTGCTAAATTCATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2383  
 Db 2221 TTTTGGATTTGCTAAATTCATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280  
 QY 2384 GTGAAGAAAGATGAACCAAGATTTTGTGCTGATGCTGAGAGAAAGCTTCAGATTCAC 2443

Db 2281 GTGAAGAAAGATGAACCAAGATTTTGTGCTGATGCTGAGAGAGAGCTTCAGATTCAC 2340  
 QY 2444 CTTCCTGGAACAGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2503  
 Db 2341 CTTCCTGGAACAGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2400  
 QY 2504 GTGAGCCAGGCTGTGAGAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2563  
 Db 2401 GTGAGCCAGGCTGTGAGAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2460  
 QY 2564 AACAGCGGCTGAAAGCCAGACCAATACAGGCTTGAACCTTACAGAGTGTGCAATCTAT 2623  
 Db 2461 AACAGCGGCTGAAAGCCAGACCAATACAGGCTTGAACCTTACAGAGTGTGCAATCTAT 2520  
 QY 2624 TACAACTGGCCAGAGTGTCAATTCGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2683  
 Db 2521 TACAACTGGCCAGAGTGTCAATTCGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2580  
 QY 2684 TTTCTGTTGGCCAGAGTATTCAGAGAGCCAAATGCTGCACTGCAACCCGCTTAC 2743  
 Db 2581 TTTCTGTTGGCCAGAGTATTCAGAGAGCCAAATGCTGCACTGCAACCCGCTTAC 2640  
 QY 2744 TACCTTAACTGCAAGAGATGACGCGCTTTCTTTTGAATGACCTTTGGAGATT 2803  
 Db 2641 TACCTTAACTGCAAGAGATGACGCGCTTTCTTTTGAATGACCTTTGGAGATT 2700  
 QY 2804 TTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTT 2863  
 Db 2701 TTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTT 2760  
 QY 2864 GGATTTAGAGATCTAGAGATTTTATTTTACAGATGCTTACACCGGCTTCTATTTAT 2923  
 Db 2761 GGATTTAGAGATCTAGAGATTTTATTTTACAGATGCTTACACCGGCTTCTATTTAT 2820  
 QY 2924 ATGTAAAAATTAAGATTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTT 2983  
 Db 2821 ATGTAAAAATTAAGATTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTT 2880  
 QY 2984 TTTTGTGTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTT 3043  
 Db 2881 TTTTGTGTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTT 2940  
 QY 3044 GAGAGTATTTGAATGATGTTTGAGATTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTT 3103  
 Db 2941 GAGAGTATTTGAATGATGTTTGAGATTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTT 3000  
 QY 3104 CTACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTT 3163  
 Db 3001 CTACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTT 3060  
 QY 3164 AAGAGGCTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTT 3223  
 Db 3061 AAGAGGCTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTT 3120  
 QY 3224 AGATTAAGATCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3283  
 Db 3121 AGATTAAGATCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3180  
 QY 3284 TTAATTTGGCAGGGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3342  
 Db 3181 TTAATTTGGCAGGGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3240  
 QY 3343 TGTCAAAAGAAACATTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTT 3402  
 Db 3241 TGTCAAAAGAAACATTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTT 3300  
 QY 3403 AACATCACTAATTAATTTTGTGTTGGAACCTGAGACATGATTTATTTGTTATTAATTAATTAATTAATTAAT 3462  
 Db 3301 AACATCACTAATTAATTTTGTGTTGGAACCTGAGACATGATTTATTTGTTATTAATTAATTAATTAATTAAT 3360  
 QY 3463 TTGATGT 3469



Db 1021 ACTGGGACCAAGAAATCCCAAGAGCACTTTAAGAAAGCCAGCGCTGTGAATGCACTTCT 1080  
 Qy 1194 TAGAATCTACAGAGAGCAATACCAACAGAGATCAACCACTTGAAGCAAGCTTGTCTTG 1253  
 Db 1081 TAGAATCTACAGAGAGCAATACCAACAGAGATCAACCACTTGAAGCAAGCTTGTCTTG 1140  
 Qy 1254 TCAGCCAGCCCAAG 1313  
 Db 1141 TCAGCCAGCCCAAG 1200  
 Qy 1314 TTCCTGAGCTCTGCTATCTTCAAGAGCTTCACTGATTAAGAGAGAGAGAGAGAGAGAG 1373  
 Db 1201 TTCCTGAGCTCTGCTATCTTCAAGAGCTTCACTGATTAAGAGAGAGAGAGAGAGAG 1260  
 Qy 1374 TGAAGAAGCTTGAAG 1433  
 Db 1261 TGAAGAAGCTTGAAG 1320  
 Qy 1434 GACTCATTTGATTAATCTTCAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1493  
 Db 1321 GACTCATTTGATTAATCTTCAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
 Qy 1494 TGAGCTTTGATTTCAACTTACTGCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1553  
 Db 1381 TGAGCTTTGATTTCAACTTACTGCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 Qy 1554 ACCAAG 1613  
 Db 1441 ACCAAG 1500  
 Qy 1614 GAGGAGCACTTAATTAAG 1673  
 Db 1501 GAGGAGCACTTAATTAAG 1560  
 Qy 1674 GAAATTTATGAG 1733  
 Db 1561 GAAATTTATGAG 1620  
 Qy 1734 GCATGCAATTAAG 1793  
 Db 1621 GCATGCAATTAAG 1680  
 Qy 1794 GAGCTTTATGAG 1853  
 Db 1681 GAGCTTTATGAG 1740  
 Qy 1854 ATCGAAG 1913  
 Db 1741 ATCGAAG 1800  
 Qy 1914 GTCAAGT 1973  
 Db 1801 GTCAAGT 1860  
 Qy 1974 AGATTGAGCTTATGAG 2033  
 Db 1861 AGATTGAGCTTATGAG 1920  
 Qy 2034 TGAAGCTGT 2093  
 Db 1921 TGAAGCTGT 1980  
 Qy 2094 CAATGAG 2153  
 Db 1981 CAATGAG 2040  
 Qy 2154 TATTTCAAGATTAAG 2213  
 Db 2041 TATTTCAAGATTAAG 2100  
 Qy 2214 TGAGGAGCTTGAATTAAG 2273  
 Db 2101 TGAGGAGCTTGAATTAAG 2160

Qy 2274 GCGTAG 2333  
 Db 2161 GCGTAG 2220  
 Qy 2334 GTCTTAAATTCATTTGTGAG 2393  
 Db 2221 GTCTTAAATTCATTTGTGAG 2280  
 Qy 2394 GAGTGAACACAG 2453  
 Db 2281 GAGTGAACACAG 2340  
 Qy 2454 CAGTTATTTGATTTGAG 2513  
 Db 2341 CAGTTATTTGATTTGAG 2400  
 Qy 2514 CTGTGAAG 2573  
 Db 2401 CTGTGAAG 2460  
 Qy 2574 TGAAGCCAG 2633  
 Db 2461 TGAAGCCAG 2520  
 Qy 2634 CAG 2636  
 Db 2521 CAG 2523

RESULT 5  
 AF264004 2328 bp mRNA linear PRI 21-AUG-2001  
 LOCUS  
 DEFINITION Homo sapiens HIWI mRNA, complete cds.  
 ACCESSION AF264004  
 VERSION AF264004.1 GI:15216446  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Sharmma, A.K., Nelson, M.C., Brandt, J.E., Wessman, M., Mulmud, N.,  
 1 (bases 1 to 2328)  
 TITLE Human CD34+ Stem Cells Express the hiwi Gene, a Human Homolog of  
 the Drosophila Gene piwi  
 JOURNAL Unpublished  
 REFERENCES 2 (bases 1 to 2328)  
 AUTHORS Sharmma, A.K. and Hoffman, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-MAY-2000) Medicine, UIC, 900 South Ashland Avenue,  
 Chicago, IL 60607, USA

FEATURES  
 source  
 1..2328  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /feature\_type="misc\_feature"  
 1..2328  
 /codon\_start=1  
 /product="HIWI"  
 /protein\_id="AAK92281.1"  
 /translation="MIFGNTNRONLDHYKESKTSQGIIVLSTNHPILTSRPMALY  
 OYHIDVNPIMBARRLRSLLFOHDLIGKCAFPQGITLFLPKRLQKTYEVSFRNG  
 EDVRLTITLLELPSTPTCLQFPYNTTFRLLKTMNQDGRNTYNPNDPIDTSHRL  
 VTPGFTTSLIQYENSIMLCTDVSHKVLRSFTVIDPFNFHQEENKFOEVSKEHL  
 GLVVLTKYNNKTYRVDIDMDQNPSTFKKADGSEVSFLERYRQYQVLEIDQPYL  
 VSPKRRRPGGTLPGPAMLIPELCYLTGLDKRNDPENVKQDLAVTRILPEQROE  
 VGRLLDYIHKNDVYORELRDWGLSFDNSLISFSGRILQTEKIHGGGTFDVPNPQFADW  
 SKETRGAPLISVPLDLNWLILYRBNYEAANSILQNLPKYTPAMGMQMRALIMLBYDD  
 RTEAYLRVLOKRYADQDIYVCLISNRKDYDAIKKYLCTDCFTPSQCVARLYGKO  
 QTVNIAIRKIALQNKCKRGELMRVDIPLKLVIMVGLDCHMDTPAGRRSLAGPVSIN  
 EGMTRWFSRCLFDGRGQELVDGLKVCIALAALRAWNSCNEVMPRIIYVRDVGDLQK

CDS  
 1..2328  
 /codon\_start=1  
 /product="HIWI"  
 /protein\_id="AAK92281.1"  
 /translation="MIFGNTNRONLDHYKESKTSQGIIVLSTNHPILTSRPMALY  
 OYHIDVNPIMBARRLRSLLFOHDLIGKCAFPQGITLFLPKRLQKTYEVSFRNG  
 EDVRLTITLLELPSTPTCLQFPYNTTFRLLKTMNQDGRNTYNPNDPIDTSHRL  
 VTPGFTTSLIQYENSIMLCTDVSHKVLRSFTVIDPFNFHQEENKFOEVSKEHL  
 GLVVLTKYNNKTYRVDIDMDQNPSTFKKADGSEVSFLERYRQYQVLEIDQPYL  
 VSPKRRRPGGTLPGPAMLIPELCYLTGLDKRNDPENVKQDLAVTRILPEQROE  
 VGRLLDYIHKNDVYORELRDWGLSFDNSLISFSGRILQTEKIHGGGTFDVPNPQFADW  
 SKETRGAPLISVPLDLNWLILYRBNYEAANSILQNLPKYTPAMGMQMRALIMLBYDD  
 RTEAYLRVLOKRYADQDIYVCLISNRKDYDAIKKYLCTDCFTPSQCVARLYGKO  
 QTVNIAIRKIALQNKCKRGELMRVDIPLKLVIMVGLDCHMDTPAGRRSLAGPVSIN  
 EGMTRWFSRCLFDGRGQELVDGLKVCIALAALRAWNSCNEVMPRIIYVRDVGDLQK

TLVNEVPOFLDCLKSIGRGNPRLTVIVKRVNTRFFFAQSGRLQNPLEPGTVIDVE  
VTRBWDYDFIVSOAVRSGSVSPHYNYVINDSLKPDHIOQLTYKLCHYIYNWPGVI  
RVPRCQVAKHLAEVLVGSIHREPNLSLNRLLYL"

BASE COUNT 716 a 496 c 520 g 596 t

ORIGIN

Query Match 66.6%; Score 2311; DB 9; Length 2328;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 2314; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy	432	TTGGTGTGAATACAGGCGAAGCCTTAGACCATGTTAAAGAAATCAAAAACAGGTTCTTTCAG	491
Db	8	TTGGTGTGAACACAGGCGAAGCCTTAGACCATGTTAAAGAAATCAAAAACAGGTTCTTTCAG	67
Qy	492	GCATTTATAGTTAAGTTAAGCATTACCATTTCCGGCTGACATCCCGTCCCGAGTGGGCT	551
Db	68	GCATTTATAGTTAAGTTAAGCATTACCATTTCCGGCTGACATCCCGTCCCGAGTGGGCT	127
Qy	552	TATATCAGTATCAATGACTATTAACCCACTGATGGAAGCCAGAGACTCCGTTTTCAGTTC	611
Db	128	TATATCAGTATCAATGACTATTAACCCACTGATGGAAGCCAGAGACTCCGTTTTCAGTTC	187
Qy	612	TTCTTTTTCACGAGAGATCTAATTTGGAAGTGTCTATGTTTTCATGATGGAACGATATAT	671
Db	188	TTCTTTTTCACGAGAGATCTAATTTGGAAGTGTCTATGTTTTCATGATGGAACGATATAT	247
Qy	672	TTTTTACTTAAAGACTACAGCAAAAGGTTACTGAAAGTCTTAAAGTAAAGCCCGGAATGGAG	731
Db	248	TTTTTACTTAAAGACTACAGCAAAAGGTTACTGAAAGTCTTAAAGTAAAGCCCGGAATGGAG	307
Qy	732	AGGATGTGAGGATTAACGATCACTTTTAAACAAATGAATCTTCAACCTCATCACCACTTGT	791
Db	308	AGGATGTGAGGATTAACGATCACTTTTAAACAAATGAATCTTCAACCTCATCACCACTTGT	367
Qy	792	TGCAGTCTATATATATTTTTCAGAGGCTTTTGAAGATTCATGAATTTGCAACAAATG	851
Db	368	TGCAGTCTATATATATTTTTCAGAGGCTTTTGAAGATTCATGAATTTGCAACAAATG	427
Qy	852	GACGAAATTTATTAACCCAAATGACCCAAATGATATTTCAAGTCAACAGTTGGTGATTT	911
Db	428	GACGAAATTTATTAACCCAAATGACCCAAATGATATTTCAAGTCAACAGTTGGTGATTT	487
Qy	912	GGCTTGCTTCACTATCTTCACTTCCATCTTCAAGTATGAAACAGCATCTGCTGCACTGAGC	971
Db	488	GGCTTGCTTCACTATCTTCCATCTTCAAGTATGAAACAGCATCTGCTGCACTGAGC	547
Qy	972	TTAGCCATAAAGTCTTTCGAAGTGAGACTGTTTGGATTTTCATGTTCAACTTTTATCATC	1031
Db	548	TTAGCCATAAAGTCTTTCGAAGTGAGACTGTTTGGATTTTCATGTTCAACTTTTATCATC	607
Qy	1032	AGACAGAAGAACATPAATTTCAAGAACAGTTTCCAAAGAAATTAAGTTAGTTAGTTGTTTC	1091
Db	608	AGACAGAAGAACATPAATTTCAAGAACAGTTTCCAAAGAAATTAAGTTAGTTAGTTGTTTC	667
Qy	1092	TTACCAAGTATTAACATTAAGACATACAGAGTGGATGATTTGCTGGGACGAGAACCCCA	1151
Db	668	TTACCAAGTATTAACATTAAGACATACAGAGTGGATGATTTGCTGGGACGAGAACCCCA	727
Qy	1152	AGAGCCTTTTAAAGAACCCGAGCGGCTCTGGGCTGAGTCTTTCATGATATCTACAGGAAGC	1211
Db	728	AGAGCCTTTTAAAGAACCCGAGCGGCTCTGAGTCTGAGTCTTTCATGATATCTACAGGAAGC	787
Qy	1212	AATACAAACAGAGATCAACCGACTTTGAAGAGCGCTGTTGGTTCAGGACCCCAAGAGAA	1271
Db	788	AATACAAACAGAGATCAACCGACTTTGAAGAGCGCTGTTGGTTCAGGACCCCAAGAGAA	847
Qy	1272	GGCGGGCCCTGGGGGAGACCTGCGAGGCTGCGATCTGCTTTCATGCTGCTGCTATC	1331
Db	848	GGCGGGCCCTGGGGGAGACCTGCGAGGCTGCGATCTGCTTTCATGCTGCTGCTATC	907
Qy	1332	TTACAGGTTCTAACTGATATAATGCTAATGATTTTAAACGTTGATGAAGACTTACCGCTTC	1391
Db	908	TTACAGGTTCTAACTGATATAATGCTAATGATTTTAAACGTTGATGAAGACTTACCGCTTC	967

Qy	1392	ATACAGACTAACTCCAGAGCAAAAGCGAGCTGAAGTGGGACGACTCATTTGATTACATTC	1451
Db	968	ATACAGACTAACTCCAGAGCAAAAGCGAGCTGAAGTGGGACGACTCATTTGATTACATTC	1027
Qy	1452	ATAAAACGATAATGTTCAAAAGGAGCTTCGAGACTGGGGTTTGAGCTTTGATTTCAACT	1511
Db	1028	ATAAAACGATAATGTTCAAAAGGAGCTTCGAGACTGGGGTTTGAGCTTTGATTTCAACT	1087
Qy	1512	TACTGTCTTCTTCAGGAAGAATTTTCCAAACAGAAAGATTTCAACAAAGTGGGAAAAACAT	1571
Db	1088	TACTGTCTTCTTCAGGAAGAATTTTCCAAACAGAAAGATTTCAACAAAGTGGGAAAAACAT	1147
Qy	1572	TTGATTAACATTCACAAATTTTGAGATTTGTCCTCAAAAGAAACAGAGCTGCACCAATTAATTA	1631
Db	1148	TTGATTAACATTCACAAATTTTGAGATTTGTCCTCAAAAGAAACAGAGCTGCACCAATTAATTA	1207
Qy	1632	GTGTTAAGCCACTAGATAACTGCTGTTGATCTATACGGGAGAAATTTATGAAGCAGCCA	1691
Db	1208	GTGTTAAGCCACTAGATAACTGCTGTTGATCTATACGGGAGAAATTTATGAAGCAGCCA	1267
Qy	1692	ATTCATTTGATACAAAATCTATTTAAAGTTTACACAGCCATGGGCATGCATAAGAAAAG	1751
Db	1268	ATTCATTTGATACAAAATCTATTTAAAGTTTACACAGCCATGGGCATGCATAAGAAAAG	1327
Qy	1752	CAATAATGATTTGAAGTGGATGACAGAACTGAAGCTTCTTAAAGTCTTACACGAAAAG	1811
Db	1328	CAATAATGATTTGAAGTGGATGACAGAACTGAAGCTTCTTAAAGTCTTACACGAAAAG	1387
Qy	1812	TCACAGCAGACACCCAGATAGTTGCTGCTGTTGTCCTGTTCAAGTCTTACGAGCAAAATACG	1871
Db	1388	TCACAGCAGACACCCAGATAGTTGCTGCTGTTGTCCTGTTCAAGTCTTACGAGCAAAATACG	1447
Qy	1872	ATGCTATTAATAAATACTGTTGATACAGATTTGCCCTTACCCCAAGTCACTGTTGGTGGCCC	1931
Db	1448	ATGCTATTAATAAATACTGTTGATACAGATTTGCCCTTACCCCAAGTCACTGTTGGTGGCCC	1507
Qy	1932	GAACTTTAGGCAACACCAACTGTCATGCTGTCCTGTTCAAAAGATTTGCCCTTACAGATGA	1991
Db	1508	GAACTTTAGGCAACACCAACTGTCATGCTGTCCTGTTCAAAAGATTTGCCCTTACAGATGA	1567
Qy	1992	ACTGCAAGATGGGAGGAGAGCTCTGGAGGTTGAGCATCTCCCTTGAAGCTCGTGAATGATG	2051
Db	1568	ACTGCAAGATGGGAGGAGAGCTCTGGAGGTTGAGCATCTCCCTTGAAGCTCGTGAATGATG	1627
Qy	2052	TTGGCATCGATTTGTTACATGACATGACAGCTGGGCGAGTCAATCGCAGGATTTCTGTTG	2111
Db	1628	TTGGCATCGATTTGTTACATGACATGACAGCTGGGCGAGTCAATCGCAGGATTTCTGTTG	1687
Qy	2112	CCAGCATCAATGAAGGGATGACCCGCTGCTGTTCTCACGCTGCATATTTTTCAGGATAGAGAC	2171
Db	1688	CCAGCATCAATGAAGGGATGACCCGCTGCTGTTCTCACGCTGCATATTTTTCAGGATAGAGAC	1747
Qy	2172	AGGAGCTGGTATAGTGGGCTCAAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2231
Db	1748	AGGAGCTGGTATAGTGGGCTCAAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1807
Qy	2232	GCAATGAGTATCATGCCAGCCGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2291
Db	1808	GCAATGAGTATCATGCCAGCCGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1867
Qy	2292	TGAAAACACTGGTGAATCTACGAAGTGGCCACAGTTTGTGATTTGTCTAAAATCTTGGTA	2351
Db	1868	TGAAAACACTGGTGAATCTACGAAGTGGCCACAGTTTGTGATTTGTCTAAAATCTTGGTA	1927
Qy	2352	GAGGTTTACAACTTATAGANTAACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	2411
Db	1928	GAGGTTTACAACTTATAGANTAACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	1987
Qy	2412	TTGCTCAGTCTGGAGGAGAGACTTTCAGAAATCCACTTTCCTGGAAACAGTTTATTTGATGAGG	2471
Db	1988	TTGCTCAGTCTGGAGGAGAGACTTTCAGAAATCCACTTTCCTGGAAACAGTTTATTTGATGAGG	2047

[illegible]

1 Kurumochi-Miyagawa, S., Kimura, T., Yomogida, K., Kuroiwa, A., Tadokoro, Y., Fujita, Y., Sato, M., Matsuda, Y. and Nakano, T. Two mouse p1w1-related genes: *m1w* and *m1l*. *Mech. Dev.* 108 (1-2), 121-133 (2001)

JOURNAL MEDLINE REFERENCE

TITLE

AUTHORS

2 (bases 1 to 3910)

Miyagawa, S.K. and Nakano, T.

Direct Submission

Submitted (21-SEP-1999) Satomi K Miyagawa, Research Institute for Microbial Diseases, Department of Molecular Cell Biology, Yamadoko 3-1, Suita, Osaka 565-0871, Japan

(E-mail: smiyagaw@biken.osaka-u.ac.jp, Tel: 81-6-6879-8362, Fax: 81-6-6879-8362)

```

FEATURES
source
Location/Qualifiers
1..3910
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="5f distal-G2 proximal"
/tissue_lib="testis"
1..3910
/gene="miwi"
156..2744
/gene="miwi"
/note="miwi family"
/codon_start=1
/product="MIMI"
/protein_id="BAA93705.1"
/db_xref="GI:7416111"
/translacion="MTGARARARARGARGOETVQHVGAANAASQPGYITPPRPOOSTPE
DIVAGRGORGWVGATSKSQEOLOISNGFOELSLARGGRRRPHLGAVTRONDLHVIA
SEKTSGGIIVKLSNHFPLSTRPOMALVOYHIDYNPMLAEARLSALLFOHEDLIG
CHAPGTILFLPKRLQHKVTEPSSORNGEHRITTLNLELPTSPCTLOFNYIIFL
RLKLTLMNLOQIGRANTYNTSPDIPIDIPNHRVYVMPGPTSLIQYENNMIMCTDSHVKVL
STVDVDFENLYQOQTEBEKFPQOVSKEILGLVLTKYNKTRVDVIDMDQPKSTFL
KSDGESEVFLVEYRKQNOEITDLKQPVLYSQPKRRRGGTLPGPAMIIPBLCVLYT
LIDKRNDEWAKMDLAHTRTLPEQOREVGLIDYIHKDQVORELRBMGSLPDSNI
LSPSRILIOSEKIHOGKTFDYNPOPADMSKEGRAPLISVPLNMLLITRRYVEE
ANSILONLFEKYPANGIOKMKAMIEVDRTETAYLALQOXYTSPQIVVCLLSNRR
DLYADIKTLCTDCTPSCQVAVRTGKOQYMAIKTALQANCKMGGELEKRVDMPE
KIAMTIDYCHDHTTAGRSIAGFPVASINEGRTMFRSCVFPDRQGEVDAGKVLQ

```

QY	164	ACAATGACTGGAGAGCCCGAGGCCAGAGCCAGAGGAAGGGCCCGGCTCAGAGACACGC	223
Db	153	AAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGCAGAGGCACGAGGTCAGAGACGGTG	212
QY	224	CAGCTGGTGGGCTCCACTCGCCAGTCAGGAACTCTGGTTATTCACGCTTAGGCTCAGCGC	283
Db	213	CAGCATGTTGGGGCTGTCTGCGAGCCAGCACTTGGGTACATCCCACTCGAGACTTAAACG	272
QY	284	CCACGACGAGAGGGGGAAATTATTTGGCCGTGACGGCAGAGAGGAA---CAGCAGAGACA	340
Db	273	TCCCCCAAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGAGTGTAGTCCGAGCC	332
QY	341	ACAGCCAATGTCACAGACATCCAGATCTCGGTGATTTCCAGAGTTATTCGATAGAGC	400
Db	333	ACATCCAAGTCAAGAACTGACAGATCTGAGTGGGTTTCAGAGACTGTCACTGGCAGAG	392
QY	401	AGAGAGGTCGTCTGATAGATTTTTCATGATCTTGGTGTGATTAACAAGCAGAACTTAGC	460
Db	393	AAGAGAGGGCGCTCGCGAGACTTCCATGACTTGGTGTGAACACGACAGAACTTAGC	452
QY	461	CATGTTAAAGATCAAAAACAGGTTCTTCAGGCATTTATATGTAAGTTAAGCATACCT	520
Db	453	CATGTCAAAGATCAAAAGCAGGCTCTCTGCAATCATTTGAACTGAGACCAACAC	512
QY	521	TTCCGGCTGACATCCCGTCCCGAGTGGGCTTATTCAGTATCACTTACATATAACCA	580
Db	513	TTCCGGCTGACCTCGCGCCCAAGTGGGCTCTGATACGATACATGACATACAAATCC	572
QY	581	CTGATGGAAGCCAGAAAGACTCCGTTGACTCTTCTTTTCAACGAGATCTAATTGGA	640
Db	573	CTGATGGAAGCCAGAAAGGCTTCGCTCCGCACTGCTCTTCACACATGAAAGCTCATTTGA	632
QY	641	AAGTGCATGCTTTGATGAGAAAGATATATTTTAACTTAAGATCAGCAAAAGTT	700
Db	633	AAGTGCATGCTTTGATGAGAAAGATATATTTTAACTTAAGATCAGCAAAAGTT	692
QY	701	ACTGAAGTTTATGTAAGACCCGGAATGAGAGATGTGAGATTAAGATCACTTTACA	760
Db	693	ACAGAAATATCAGTCAGACTCGAAATGGGGAACAGTAGGATCACATCACTCCGACC	752
QY	761	AATGAATTCACATCATCAACCACTTGTGTGACGTTCTATATATATTTTTCGAGAG	820
Db	753	AACGAGCTCCCGCCACCTCGCCACCTGCGCACTTATTAACATCATCTTCAGAGAG	812
QY	821	CTTTGAAATCATGAATTTGCAACAATTTGACAGAAATTTATTAACCAATGACCA	880
Db	813	CTCTTGAATCATGAATTTGCAACAATTTGACAGAAATTTATTAACATCAAGTGACCG	872
QY	881	ATTGATATTTCAAGTCACAGGTTGTGATTTGGCTTGCTTCACTTCCATCTTCAG	940
Db	873	ATTGATATTTCAAAACACAGGTTGTGATCTGGCCCGGCTTCAACCACTTCATCTTCAG	932
QY	941	TATGAAACAGACATCATGCTGCACTGACGTGACGTTAGCATTAAGTCTTGAAGTGAAGT	1000
Db	933	TATGAAACAGACATCATGCTGCACTGACGTCGACGCAAGGTGCTCGCACGAGACT	992
QY	1001	GTTTGGATTTCAATGTTCAACTTTATATCATCAGACAGAAACATTAATTTCAAGACAA	1060
Db	993	GTCCTGACTTCAATGTTCAATCATATATCCAGACAGAGAGACACAAGTTCCAGGAGCAA	1052
QY	1061	GTTTCAAGAAATATATAGTTTATGTTGTTCTTACCAAGTATATACATTAAACATACAG	1120

Db 1053 GTGTGAAGGAGCTCATAGGCTCATCGTTCTCACCAAGTACAAATACAGAGCTACCGG 1112  
 Qy 1121 GTGATGATATTTGACTGGGACGAGATCCAGAGGACCTTTTAAAGAGCCGAGCGGTCT 1180  
 Db 1113 GTGATGACATTTGACTGGGACGAGATCCAGAGGACCTTTTAAAGAGCCGAGCGGTCTG 1172  
 Qy 1181 GGGGTGAGCTTTTGTAGAAATCTACAGGAAGCAATACAAACAGAGATCACCGACTTTGAAG 1240  
 Db 1173 GAGGTGAGCTTTTGTGGAGTACTACAGGAAGCAATACAAACAGAGATCACCGACTTTGAAG 1232  
 Qy 1241 CAGCTGTCTTTGTGTGACGACCCCAAGAGAGCGGGGCGCTTGGGGGACACTGCCAGGG 1300  
 Db 1233 CAGCGGTGTGTGTGAGCCCAACCAAGCGAGAGAGCGGGCGCGCTTGGCTGGC 1292  
 Qy 1301 CTGCGATGCTCATCTCTGAGCTCTGTCTATCTTACAGGTCTAACTGATAAATGCGTAAAT 1360  
 Db 1293 CCAGCTATGCTCATCTCTGAGCTCTGTCTATCTTACAGGTCTAACTGATAAATGCGTAAAT 1352  
 Qy 1361 GATTTTAACTGTGATGAAGACTTAGCGTTTCATACAGACTACTCCAGAGCAAGCGAG 1420  
 Db 1353 GATTTTCAATGTGATGAAGACTTAGCGTTTCATACAGACTACTCCAGAGCAAGCGAG 1412  
 Qy 1421 CGTGAAGTGGGACGACTCATTTGATTAATTAATTAATAAAACGATTAATTTCAAAGGGAGCTT 1480  
 Db 1413 CGGAGGTGGGCGCTCATCTGACTATCATCAAGATGACAAATGTGAGAGAGACTT 1472  
 Qy 1481 CGAGACTGGGTTTGTAGCTTTGATTTCAACTTCACTTCTTCTCAGAGAGAAATTTGAAA 1540  
 Db 1473 CGAGACTGGGCGCTGAGCTTTGACTTCAAACTTGTCTCTCTGGAAGAAATCTTACAA 1532  
 Qy 1541 ACAGAAAGAAATTCACCAAGGTGGAAACAAATTTGATTAATTAATAAACTTTCAGAGTTGG 1600  
 Db 1533 TCTGAGAAATTCACCAAGGTGGAAACAAATTTGATTAATTAATAAACTTTCAGAGTTGG 1592  
 Qy 1601 TCCAAAGAAAACAGAGGTGACCACTTAATTAATTTAGTGTAAAGCCACTAGATACTGGCTGTG 1660  
 Db 1593 TCCAAAGAAAACAGAGGTGACCACTTAATTAATTTAGTGTAAAGCCACTAGATACTGGCTGTG 1652  
 Qy 1661 ATCTATACGGAAGAAATTAAGAGAGCAAGCAATTAATTAATTAATAAACTTTCAGAGTT 1720  
 Db 1653 ATCTATACGGAAGAAATTAAGAGAGCAAGCAATTAATTAATTAATAAACTTTCAGAGTT 1712  
 Qy 1721 ACACGAGCAATGGGCTGCAAAATGAGAAAGCAATTAATTAATTAATAAACTTTCAGAGTT 1780  
 Db 1713 ACTCCAGCAATGGGCTGCAAAATGAGAAAGCAATTAATTAATTAATAAACTTTCAGAGTT 1772  
 Qy 1781 GAAGCTTACTTAAGTCTTACAGCAAAAGGTTCACAGCAGACACCCAGATAGTTGTCTGT 1840  
 Db 1773 GAAGCTTACTTAAGTCTTACAGCAGCAAGGTTCACAGCAGACACCCAGATAGTTGTCTGT 1832  
 Qy 1841 CTGTTGTCAGTAATCGGAAGGCAAAATGAGATGCTTATTAATAAACTTTCAGAGTT 1900  
 Db 1833 CTGTTGTCAGTAATCGGAAGGCAAAATGAGATGCTTATTAATAAACTTTCAGAGTT 1892  
 Qy 1901 TGCCCTTACCCCAAGTCAAGTGTGTGTGGCGGCAACCTTGAAGCAACAGCAAACTGTCTATG 1960  
 Db 1893 TGCCCTTACCCCAAGTCAAGTGTGTGTGGCGGCAACCTTGAAGCAACAGCAAACTGTCTATG 1952  
 Qy 1961 GCAATTTGCTACAAAGATTCGCTTACAGATGAACCTGCAAGATGGGAGGAGCTCTGGAGG 2020  
 Db 1953 GCAATTTGCTACCAAGATTCGCTTACAGATGAACCTGCAAGATGGGAGGAGCTCTGGAGG 2012  
 Qy 2021 GTGGACATCCCTTGAAGTCTGTGATGATGCTTGGCATGATGTTTACCATGACATGACA 2080  
 Db 2013 GTGGACATCCCTTGAAGTCTGTGATGATGCTTGGCATGATGTTTACCATGACATGACA 2072  
 Qy 2081 GTGGCGGAGGTCAATTCGAGGATTTGTGTCAGCATCAATGAAGGGATGACCGCTGG 2140  
 Db 2073 GCTGGCGGAGGTCCATTCGAGGATTTGTGTCAGCATCAATGAAGGGATGACCGCTGG 2132  
 Qy 2141 TTCTACGCTGCATATTTTCCAGATGAGAGCAAGGAGCTGGTATAGTGGGCTCAAACTGTCG 2200  
 Db 2133 TTCTCCCGTGGCTCTTTTCCAGGACCGGCGGAGGAGCTGGTGGATGGTCTCAAGGTGTGC 2192

Qy 2201 CTGCAAGCGGCTCTGAGGGCTTGGAAATAGCTGCAATGAGTACATGCCCAGCCGATCATC 2260  
 Db 2193 TTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGTTGTCATC 2252  
 Qy 2261 GTGTACCCGATGCGGTAGGAGACGGCCAGCTGAAACACATGTTGAACTACGAAGTGGCA 2320  
 Db 2253 GTGTACCCGATGCGGTAGGAGACGGCCAGCTGAAAGACCTTGGTCAATATGAGGTGCCA 2312  
 Qy 2321 CAGTTTTTGGATTTGCTAAAAATCCATTTGTTAGAGTTTCAACACCTAGANTAAACGGTAATT 2380  
 Db 2313 CAGTTTCTTAGATTGCTTCAAGTCACTGCGGAGAGTTTCAACACCAAGACTGACTGTAAATC 2372  
 Qy 2381 GTGTGAAGAAAAGAGTGAACACACAGATTTTTTGTCTCAGTCTGGAGGAAGACTTCAAGAT 2440  
 Db 2373 GTGTGAAGAAAAGAGTGAACACACAGATTTTTTGTCTCAGTCTGGGGAAGACTTCAAGAT 2432  
 Qy 2441 CCACCTTCTCTGAAACAGTTATTTGATGTAGAGTTTACACAGCAGCAATGTTATGACTTTTTT 2500  
 Db 2433 CCTCTTCCAGGAGCAGTCACTGATGTGAAAGTACCAGACCAAGTGTATGACTTTTTTTC 2492  
 Qy 2501 ATCTGTAGCCAGGCTGTGAGAAAGTGTGTGTTTCTCCACACATTAACAATGTCATCTAT 2560  
 Db 2493 ATCTGTAGCCAGGCTGTGAGAAAGTGTGTGTTTCTCCACACATTAACAATGTCATCTAT 2552  
 Qy 2561 GACAAAGCGGCTTGAAGCCAGACCAATACAGCGCTTGACCTTCAAGCTGTGCCAATC 2620  
 Db 2553 GACAGCAGTGGCTTGAAGCCAGACCAATACAGCGCTTGACCTTCAAGCTGTGCCAATC 2612  
 Qy 2621 TATTACAACTGGCCAGGCTGTCATTTGTTTCTTCTGCTTGGCAGTACGCCCAAGCTG 2680  
 Db 2613 TACTATTAATTTGGCTTGGAGTCACTCCGAGTCCCTGACCTTGGCAGTATGCACAAAGCTG 2672  
 Qy 2681 GCCTTTCTTGTGTGCCAGAGTATTTCAGAGAGCCAAATCTGTCACTGTCAAAACCGCTT 2740  
 Db 2673 GCCTTTCTTGTGTGCCAGAGTATTTCAGAGAGCCAAATCTGTCACTGTCAAAACCGCTT 2732  
 Qy 2741 TACTACTCTTAACCTGCAG 2759  
 Db 2733 TACTACTCTTAACCTGCAG 2751

RESULT 7  
 AF438405 Mus musculus MIWI mRNA, complete cds. linear ROD 13-NOV-2001  
 LOCUS AF438405  
 DEFINITION Mus musculus MIWI mRNA, complete cds.  
 ACCESSION AF438405  
 VERSION AF438405.1 GI:16905060  
 KEYWORDS  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 Deng, W. and Lin, H.  
 1 (bases 1 to 4064)  
 Miwi, a murine homolog of piwi, encodes a cytoplasmic protein  
 essential for spermatogenesis  
 Unpublished  
 2 (bases 1 to 4064)  
 Deng, W. and Lin, H.  
 Direct Submission  
 Submitted (22-OCT-2001) Department of Cell Biology, Duke  
 University, Room 412, Nannaline Duke Building, Durham, NC 27710, USA  
 Location/Qualifiers

1. .4064  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="5"  
 /tissue\_type="testis"  
 /note="piwi-like protein"  
 /codon\_start=1  
 /product="MIWI"

CDS





Db 1988 GGCATTCGCCACCAAGATCGCCCTGCAGATGAATCAAGATGGGAGCGAGCTTCGGCG 2047  
 Qy 2021 GTGGACATCCCTGAAGCTCGTGATGTCGTGGATCGATGTTGATGTTACCATGACATGACA 2080  
 Db 2048 GTGGACATCCCTGAAGCTCGATGATGTCGTGGATCGATGTTACCATGACATGACA 2107  
 Qy 2081 GCTGGCGGAGGTCAATCGCAGGATTTGTCGACGATCAATGAAGGATGACCGGTGG 2140  
 Db 2108 GCTGGCGGAGGTCCATCGCAGGATTCGTGGCGGAGGATGAAGGATGACCGGTGG 2167  
 Qy 2141 TTCATCGCTGCATATTTCCAGATGAGGACAGGAGCTGGTAGTGGGCTCAAGTCTGC 2200  
 Db 2168 TTCCTCGCTGCTCTTTCCAGGACCGCGGAGGAGCTGGTAGTGGTCTCAAGTGTGC 2227  
 Qy 2201 CTGCAAGCGGCTCTGAGGCTTTGGAATAGCTCAATGAGTACATCCCGGCGGATCATC 2260  
 Db 2228 TTGCAAGCTGCTCTGAGGCTTTGAGTGGCTGCTCAATGATATCATGCGCGGTCTCATC 2287  
 Qy 2261 GTGTACCGGATGGGTAGGACCGCCAGCTGGAACACATGGTGAACTACGAAAGTGCCA 2320  
 Db 2288 GTGTACCGGAGCGGTGTGGGGGACGGGAGCTGGAAGACCTGCTCAATATGAGGTCCA 2347  
 Qy 2321 CAGTTTTTGGATGCTAAATCCATTGCTAGAGGTTACACCTAGANTACGGTAATT 2380  
 Db 2348 CAGTTTCTAGATTCGCTCAAGTCAGTCGGGAGAGGTTACACCCAGAGCTGACTGTAATC 2407  
 Qy 2381 GTGGTGAAGAAAGAGTGAACACGAGATTTTGTCTCAGTCTGGAGGAGAGCTTCAGAAAT 2440  
 Db 2408 GTGGTGAAGAGCGGTGCAATGCCAGTTTGTCTCAGTCTGGGGAAGAGCTTCAGAAAC 2467  
 Qy 2441 CCATTCCTTGGACAGTTATGATGATGAGGTTACAGACAGAAATGGTATGACTTTTTC 2500  
 Db 2468 CTTCTTCCAGGACAGTCAATGATGGAAGTCAACAGACAGAGTGGTATGACTTTTTC 2527  
 Qy 2501 ATCGTACGACGCTGTGAGAGTGTGAGTGTCTTCCACACATTTACAAATGTCATCTAT 2560  
 Db 2528 ATCGTACGACGCTGTGAGAGTGTGAGTGTCTTCCACACATTTACAAATGTCATCTAT 2587  
 Qy 2561 GACACAGCGGCTGAAGCCAGACACATACAGCGCTTGACCTACAAAGCTGTGCCACATC 2620  
 Db 2588 GACAGAGTGGCTGAAGCCGACACATCCAGCGCTGACATACAAAGCTGTGCCACATG 2647  
 Qy 2621 TATTACAACTGGCAGGTGATTCGTTGTTCTGCTTCCAGTACGACGACGACGACGCTG 2680  
 Db 2648 TACTATAATGGCTGGAGTCAATCCAGTCCCTGACGCTTGGCAGTATGACACAAAGCTG 2707  
 Qy 2681 GCTTTTCTGTTGGCCAGAGTATTCACAGAGGCGCAATCTGCTCACTGTCAAAACCGCTT 2740  
 Db 2708 GCCTTCTCGTGGGCCAGAGTATTCACAGAGGCGCAATCTTCTCTGTTCCAAACCGCTC 2767  
 Qy 2741 TACTACTCTTAACCTGCAG 2759  
 Db 2768 TACTACTCTTAACAGAG 2786

RESULT 8  
 AF336369 3273 bp mRNA linear VRT 02-JAN-2002  
 LOCUS  
 DEFINITION  
 AF336369  
 ACCESSION  
 VERSION  
 AF336369.1 GI:18033486  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Danio rerio.  
 Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.  
 1 (bases 1 to 3273)  
 Weeratne,S.D., Gong, Z. and Tan, C.-H.  
 Cloning and characterization of zebrafish homolog of piwi,  
 essential for germ-line stem cell self-renewal  
 Unpublished

# REFERENCE AUTHORS TITLE JOURNAL

2 (bases 1 to 3273)  
 Weeratne,S.D., Gong, Z. and Tan, C.-H.  
 Direct Submission  
 Submitted (15-JAN-2001) Department of Biological Sciences, National  
 University of Singapore, 10 Kent Ridge Crescent, Singapore 119260,  
 Singapore

## FEATURES

### Location/Qualifiers

1..3273  
 /organism="Danio rerio"  
 /db\_xref="taxon:7955"  
 64..2640  
 /codon\_start=1  
 /product="piwi protein"  
 /protein\_id="AA157170.1"  
 /db\_xref="GI:18033486"

## CDS

/translation="MTGRARASRGGRGPAPAGPAPVPSQEAAPVVPSTPBGQL  
 VGRGKDPAPGAMSEAMLOISAGQVQYKHIGERRRDFDSDGVTHTQLMEHVKEK  
 TGVGTATLRLANFRLSRLPMWALYQHVVDYKPMESRRLLSALLPOEHTLGAHT  
 FDGALLFLPNKLRNAETVLCSETRNGEKEVITVLTNELPPSSPCLQVSHILPRIL  
 RLNMQQIGRHYNPDPDPFPIQHLRTTWPGFMTTLQYESSIMLCSVSHKVRSET  
 VLDFMSLRQCGDORFEACTKELVGLIILTKNNKTYRIDDDIANDHTPNTFKGD  
 TEISFKNYFKSQYGLDITDGNQVLLSHVKRLGSPGPPGAPMLVPEFCYLTGLTDK  
 MEADINIMKDLASHTLSPEQREGRINRLISININGVDVQNELTWTGLSFENKLSLN  
 GRVLSERIIOGGRAFEXNPMWTDWSEKMRGLPLISCSLDNLMFLYFRNADVAQSL  
 LOTLANKVSGPMGIRMORAVMTIEDROESLLRALQONVARETOMVVILPTNRDKYD  
 CVKYLVDCTPTPSQCVSRSTISKPOLMTVATKALOMCKMGGELWSVEIPLRLM  
 IVGIDCIHDTHAKRSIGAMVASLNMQMSRFKCVLQNRQEIIDALKGSLQALKA  
 YLKNNSLPRIIVYRDGDMQSVVDYEVFPQIMQSIKTMGQDDEPKLSVVVVKRR  
 ISSRFARIDGKIANDPPTVIDTEPEWYDFIVSOAVRFGCAVTHYNNVPDNS  
 GLKPDHMQRLTYKLCHMYNNQGIIVRVPAPQYAHKLAFLVQGSIHKEPNMLDDFLY  
 YL"

BASE COUNT 922 a 717 c 787 g 847 t

Query Match 30.7%; Score 1067.2; DB 5; Length 3273;  
 Best Local Similarity 63.3%; Pred. No. 6.6e-239;  
 Matches 1652; Conservative 0; Mismatches 956; Indels 3; Gaps 1;

Qy 153 AGAAATAGAAAACATGACTGGGAGAGCCGAGCCAGAGCCAGAGAGAGGCGCGGTC 212  
 Db 44 AGCATTTGTTAAAGACCGGTATGACAGAGCAGCAAGAGATCGAGGGGCGAGGTC 103  
 Qy 213 AGGAGACAGCGCAGCTGGTGGGCTCCACTGCCAGTCCAGCAACCTGTTATATTCAGCCTA 272  
 Db 104 GTGGTCAGGAGCTCTGCTCTCCGGGCGCAACCACTGTATCTCAAGAGGTCGCAAGC 163  
 Qy 273 GGCTCAGCGCCACACAGACAGAGGGGAAATATTTCGGCGGTGAGCGGAGAGAGAACAG 332  
 Db 164 CTGTTGTGTCCTCCGCTCTGAAGGACAACTGGTGGGCGAGAGGAGACAGAAACCTGCGAC 223  
 Qy 333 CAGGAGAACAGCAAGTCAAGAGGATCCAGATATCTGCTGGATTTTCAGGAGTTATCGN 392  
 Db 224 CTGGAGCCATGTCTGAAGAGCGATGCTCAGATTTCTGCTGGCTTTTCAGAGGTTGAAGA 283  
 Qy 393 TAGCAGAGAGAGAGGTCGTCGTAGAGATTTTCATGATCTTGTGTGTAATACAAAGGCAGA 452  
 Db 284 TCGGGAGAGAGGTGAGCGCAGACGGGATTTCCAGATCTCTGCTGTCATCTCGTCAC 343  
 Qy 453 ACCTAGACCATGTTAAAGATCAAAAACAGGTTCTTCAGGCATATATAGTAAGTTAAGCA 512  
 Db 344 TTATGGAACACGTGAAGAAATCCAAAACAGGTGTTCTTGGGACTGCCATTTGAGCTGAGGG 403  
 Qy 513 CTAACCATTTCCGCTGACATCCCGTCCCGAGGCGCTTATATCATGATTCATGATTCATTTGACT 572  
 Db 404 CCAATTTTATGCGTCTGCTTTCAGCTTCTATGTTGGCTTTGTTATCATGATTCATGCTGACT 463  
 Qy 573 ATAACCCACTGATGGAAGCAGAGAGACTCCGTTTTCAGCTCTCTTTTTCACACGAAAGATC 632  
 Db 464 ACAGCCACCAATGGAGTCCAGACGACTGCTGCTTCCGCTTTGCTTTTCCAGCAGAGGAGA 523  
 Qy 633 TAATTGAAAGTGTGATGCTTTTGTATGGAACGATATATTTTACCTAAAGACTACAGC 692  
 Db 524 CTCTGGGGAAGACACACACCTTTTGTATGGAGCCATTCTCTTTCTCCCAACAACTGCGCA 583



693 AAAAGGTTAAGTATTTAGTAAAGCCGGAATGAGAGATGAGATTAAGATCA 752  
584 ACCCTGAGACAGTCTGTCGAGTGAAGCAAGAAATGAGAGAGTGAATCAGTGA 643  
753 CTTTAAATATGAACTTCCACTACATCAACAATTGTTGAGCTTCTATTAATATT 812  
644 CACTGACATATGAGTCCGCGCTCTCTCTGATGCTGAGCTTTCACATCTCTCT 703  
813 TCAGAGAGCTTTTGAATATCATGAAATTTGCAACAAATTTGAAGAAATTTAAT 872  
704 TCAGAGAAATTCCTAAGAAATCTAAGAAATGAGAGATGAGAGCTTACTACAT 763  
873 ATGACCCATTTGATATTCAGAGTCAAGTGTGATTTGAGCTGAGCTTCACTTCA 932  
764 ATGACCCATTTGATATTCAGAGTCAAGTGTGATTTGAGCTGAGCTTCACTTCA 923  
933 TCTTCAATATGAAAGAGATCATGCTGTCATGAGCTTATGCTTATGCTTCAAT 992  
824 TCTTCAATATGAAAGAGATCATGCTGTCATGAGCTTATGCTTATGCTTCAAT 883  
993 GTGAGAGCTGTTTGAATTCATGTTCACTTTATCATCAGACAGAGAACTAATTT 1052  
884 GTGAGAGCTGTTTGAATTCATGTTCACTTTATCATCAGACAGAGAACTAATTT 943  
1053 AAGAACAAATTTCCAAAGAAATTAATAGGTTTATGTTTCTTACAGATTAATTA 1112  
944 CTGAAGCTGACAGAGAGAGCTTGTGGCTGATCATTTGACGAGTACACACAGAA 1003  
1113 CATACAGATGATGATATTTGATCTGAGACAGAAATCCCAAGAGACCTTTAAGAA 1172  
1004 CATACAGATGATGATATTTGATCTGAGACAGAAATCCCAAGAGACCTTTAAGAA 1063  
1173 ACGGCTCTGGGCTCAGCTTCTTGAATCTAAGAAACAAATCAAGAAAGTCAACG 1232  
1064 A---TACAGAAATTTCAATCAAGAACTACTTCAATTCGAATTTGATTAACAG 1120  
1233 ACTTGAAGAGAGCTGCTGCTGTCAGACAGAGAGAGAGAGAGAGAGAGAGAGAG 1292  
1121 ATGGCAACAGAGTCTTCTGCTTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 1180  
1293 TGGCAG 1352  
1181 CACCGGAG 1240  
1353 TGGGATATGATTTTAAAG 1412  
1241 TGGGATATGATTTTAAAG 1300  
1413 AAG 1472  
1301 AG 1360  
1473 GGGAGCTTCAAG 1532  
1361 ATGAG 1420  
1533 TTTTGAAG 1480  
1421 TTTTGGCTCTGAG 1540  
1593 CAGATTTGAG 1652  
1481 CAGATTTGAG 1540  
1653 GGGTGTGATCTATACAG 1712  
1541 GGGTGTGATCTATACAG 1600  
1713 TTTAAGTATACAG 1772  
1601 ACAG 1660

1773 ACAG 1832  
1661 ATGCTGAG 1720  
1833 TTTGCTGATCTGTTCAAG 1892  
1721 TGGTGTGATCTATACAG 1780  
1893 GTACAG 1952  
1781 GTGAG 1840  
1953 CTGCTGATGAG 1900  
1841 CACTTATGAG 2072  
2013 TCTGAG 2072  
1901 TGTGAG 1960  
2073 ACATGAG 2132  
1961 ATACCGCTGAG 2020  
2133 CCGCTGATCTGAG 2192  
2021 CAG 2080  
2193 AAGTGTGCTGAG 2252  
2081 AAGGATCATTTGAG 2140  
2253 GGATCATGCTGAG 2312  
2141 GCATCATGCTGAG 2200  
2313 AAGTGTGCTGAG 2372  
2201 AAGTGTGCTGAG 2260  
2373 CGGTATGCTGAG 2432  
2261 CAGTATGCTGAG 2320  
2443 TTTGAG 2492  
2321 TTTGAG 2380  
2493 ACTTTTATGAG 2552  
2381 ATTCTTATGAG 2440  
2553 TCATCTATGAG 2612  
2441 TGGTGTGAG 2500  
2613 GGCACATCTATGAG 2560  
2501 GGCACATCTATGAG 2632  
2673 ACAG 2732  
2561 ACAATTTGAG 2620  
2733 ACGGCTTATGAG 2763  
2621 ACTTCTGATGAG 2651

RESULT 9  
AC025837/c AC025837 DNA linear HTG 06-MAY-2001  
LOCUS Homo sapiens clone Rpl1-200K12, WORKING DRAFT SEQUENCE, 29  
DEFINITION

```

unordered pieces.
ACCESSION AC025837
VERSION AC025837.3 GI:8072609
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 193697)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome, clone RP11-200K12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 193697)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,
Boguslavsky,L., Bouckgaier,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galegani,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessier,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced gi:7342026.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5437
Center clone name: 200 K.12
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Assembly: Dye-terminator Big Dye; 100% of reads
Consensus quality: Phrap; version 0.960731
Consensus quality: 174594 bases at least Q40
Consensus quality: 183716 bases at least Q30
Consensus quality: 187939 bases at least Q20
Insert size: 183000; agarose-fp
Quality coverage: 3.9 in Q20 bases; agarose-fp
Quality coverage: 3.7 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1628: contig of 1628 bp in length
* 1629 1728: gap of 100 bp
* 1729 2962: contig of 1234 bp in length
* 2963 3062: gap of 100 bp
*

```

```

* 3063 4989: contig of 1927 bp in length
* 4990 5089: gap of 100 bp
* 5090 6666: contig of 1577 bp in length
* 6667 6766: gap of 100 bp
* 6767 8667: contig of 1901 bp in length
* 8668 8767: gap of 100 bp
* 8768 11259: contig of 2492 bp in length
* 11260 11359: gap of 100 bp
* 11360 13657: contig of 2198 bp in length
* 13658 17173: gap of 100 bp
* 17174 17273: contig of 3516 bp in length
* 17274 20060: contig of 2787 bp in length
* 20061 20160: gap of 100 bp
* 20161 23843: contig of 3683 bp in length
* 23844 23943: gap of 100 bp
* 23944 27203: contig of 3260 bp in length
* 27204 27303: gap of 100 bp
* 27304 31592: contig of 4289 bp in length
* 31593 31692: gap of 100 bp
* 31693 36740: contig of 5048 bp in length
* 36741 36840: gap of 100 bp
* 36841 42197: contig of 5357 bp in length
* 42198 42297: gap of 100 bp
* 42298 47652: contig of 5355 bp in length
* 47653 47752: gap of 100 bp
* 47753 52307: contig of 4555 bp in length
* 52308 52407: gap of 100 bp
* 52408 58606: contig of 6199 bp in length
* 58607 58706: gap of 100 bp
* 58707 63584: contig of 6878 bp in length
* 63585 65684: gap of 100 bp
* 65685 72492: contig of 6808 bp in length
* 72493 72592: gap of 100 bp
* 72593 78756: contig of 6164 bp in length
* 78757 78856: gap of 100 bp
* 78857 85474: contig of 6618 bp in length
* 85475 85574: gap of 100 bp
* 85575 91951: contig of 6377 bp in length
* 91952 92051: gap of 100 bp
* 92052 99044: contig of 6993 bp in length
* 99045 99144: gap of 100 bp
* 99145 104884: contig of 5740 bp in length
* 104885 104984: gap of 100 bp
* 104985 112130: contig of 7146 bp in length
* 112131 112230: gap of 100 bp
* 112231 118017: contig of 5787 bp in length
* 118018 118117: gap of 100 bp
* 118118 126319: contig of 8202 bp in length
* 126320 126419: gap of 100 bp
* 126420 137651: contig of 11232 bp in length
* 137652 137751: gap of 100 bp
* 137752 193697: contig of 55946 bp in length.
FEATURES
source
Location/Qualifiers
1..193697
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-200K12"
/clone_lib="RP11-11 Human Male BAC"
1..1628
/note="assembly_fragment"
1729..2962
/note="assembly_fragment"
3063..4989
/note="assembly_fragment"
5090..6666
/note="assembly_fragment"
6767..8667
/note="assembly_fragment"
8768..11259
/note="assembly_fragment"
11360..13657
/note="assembly_fragment"

```

Db	Accession	Version	Keywords	Source	Organism	Reference	Authors
Db 138722	CTAGCATTTTATTTCTAGACATTCCTATTTCACCGCGCTCTTATTTTATACGTAAATAATA						
Qy	2936	AGATTTTATATTTTATATCTTCTGTTCTCATAGATATTTTGTGACATTTTGTGTTAT					
Db 138662	AGATTTTATATTTTATATCTTCTGTTCTCATAGATATTTTGTGACATTTTGTGTTAT						
Qy	2996	TTTGAGAAAATGTGATATAGATATCTTGTTGTATATTAACAAGATCTCTGAGATATTTGA					
Db 138602	TTTGAGAAAATGTGATATAGATATCTTGTTGTATATTAACAAGATCTCTCTGAGATTTTGA						
Qy	3056	AATGTGTTTGGAGATTTTACTTAAACGTAATCTTTCAGAGTGTAGGCAAGTCTTATTATTAAC					
Db 138542	AATGTGTTTGGAGATTTTACTTAAACGTAATCTTTCAGAGTGTAGGCAAGTCTTATTATTAAC						
Qy	3116	CTATATTAATCTTATTTTGTGATATCTGTTTGAATTTTAAAGAGATTAAGAGCGCTTAA					
Db 138482	CTATATTAATCTTATTTTGTGATATCTGTTTGAATTTTAAAGAGATTAAGAGCGCTTAA						
Qy	3176	GTAAGATGCTACATCAACATAGATGAGGAGTTTCAGCTCATATCTTAAAGATTAAGATTA					
Db 138422	GTAAGATGCTACATCAACATAGATGAGGAGTTTCAGCTCATATCTTAAAGATTAAGATTA						
Qy	3236	CTATATTAATCACTTATACCAAGATTCAGAGAGAAATATGCTGATTTTATTTTGGCAGG					
Db 138362	CTATATTAATCACTTATACCAAGATTCAGAGAGAAATATGCTGATTTTATTTTGGCAGG						
Qy	3296	GGGGCTAGGTTGTATAGGAGGTAAATAAACAATGGAAATTTTAAATTTGTCCTAAGAAAC					
Db 138302	GGGGCTAGGTTGTATAGGAGGTAAATAAACAATGGAAATTTTAAATTTGTCCTAAGAAAC						
Qy	3356	ATTTTAAGACTCTTTTACAAAAGGCAATGATGTAATCTGATTTTAAATCACTATATTT					
Db 138242	ATTTTAAGACTCTTTTACAAAAGGCAATGATGTAATCTGATTTTAAATCACTATATTT						
Qy	3416	ATTTTGTTTTGAAGACTGGACATGATTTCTATTTTGTATTAATAATTAATTTGATGT					
Db 138182	ATTTTGTTTTGAAGACTGGACATGATTTCTATTTTGTATTAATAATTAATTTGATGT						
RESULT 10							
AC127071							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
AUTHORS							

REFERENCE  
AUTHORS  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 3006)  
Strausberg, R.

TITLE  
JOURNAL

Direct Submission  
Submitted (03-JUN-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

REMARK  
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contract: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Miklos Palakovic, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [medepax1.stanford.edu](mailto:medepax1.stanford.edu)  
Dickson, M., Schmitz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRAC Plate: 48 Row: B Column: 15  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Similarity but not  
identity to protein.

FEATURES  
source

Location/Qualifiers  
1. 3006  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:33296 IMAGE:5271278"  
/issue\_type="Testis"  
/clone\_lib="NIH MGC\_97"  
/lab\_host="DH10B"  
/note="Vector: pBluescript"  
4. 2562  
/codon\_start=1  
/product="Similar to p1w1 like homolog 1 (Drosophila)"  
/protein\_id="AAH31060.1"  
/db\_xref="GI:21410561"  
/translation="MSGARVAKRGIARSPATSEVGRIOQLPSRVSVDLSNNBASSNS  
GFLGSRISTNDKVGISDAGSTFEMRGVNDKDFDLSICRELIARVNRKCTGGS  
GIPVAVLNLPHLDIPDPMOLYOHVHYIPDLASRLRILALSHSLIYNDNRKCTGGS  
ILFLSKLEKVELSSETQGETIKWITILKRLPSSVPCIQVNIILFRKLKLS  
MYOIGNFNPSEPELIPQHKLSLMPGALISVYFEKRLFSADVSKYLVNLTVEIF  
MTACORTGLSCFTQCEKQILGLIVLRNNRYSDDLDMGKPTHTFOKDGTEI  
TYDYKQOYDITVSDNQPMLVSLKRRDMSBAQLAHLELCLEIGLIDQATSD  
FOLKAVAKRSLSPSGRQOLARLVNDIQRNTARFELHMGHRSQSLSTRIRVP  
SEKILMODHICOPVSAADMSKDIRCKILNOSNTWLLICSDTEYVASSPNCIR  
VASMGPNDVYPRKIKVQENPAPAFVRAIQYVDVQVLMCISDNKTYDSIKTL  
SDCPVPSQCVLARTLNKQMMMSIATKLAOMTCIKGELMAVEILKLMVGIIDV  
KDALSDVNVGCVASVNPRTIRWPSCLIORMTVDADCLAKFMGALNKMYKKNH  
DLPARLIVNAGVGDQKTLLEYVPLISVASSNTSRLSVIVYKCKPRF  
TEMRVQNPPLIGTIVDSATRNEMWDFILISQVACRTVSPYVNIYDNGKPKDH  
MORLFECLHYNNMPGIVSPAPQYKRLFTVAOSIHKEPSELNHLFLY"

CDS

BASE COUNT  
ORIGIN

Query Match 19.6%; Score 679.2; DB 9; Length 3006;  
Best Local Similarity 56.6%; Pred. No. 4.7e-148;  
Matches 1320; Conservative 0; Mismatches 1005; Indels 9; Gaps 3;

419 GATTTCATGATCTGTGTAATCAAGAGCACTGACCACTGTTAAAGAACTCAAA 478  
238 GACTTATGATTTAGTATCTGTACCAAGAAAATTGGCACTGTGAAGAAATTGTA 297  
479 ACAGGTTCTTCAGGCACTTATAGTAAGTTAGCACTTACCACTTCCGCTGACATCCGT 538  
298 ACAGGTTCCAGTGAATACCTGTGAATCTGTACCACTTCTTAACTTGAATTTCCC 357  
539 CCCAGTGGGCTTATATCACTGATCACTTAACTTAACTGATGAGGAGCAAGAA 598  
358 CAAGCTGGCACTTATCACTGATCACTGATGATGATGATGATGATGATGATGATG 417

599 CTCCTGTCAGCTCTCTTTTTCACACAGAGATCTAATGAAAGTGTCTTTGAT 658  
418 CTGAGATTTGCTTCTTATATGATGTAATCTTCAACAAAGCAAAAGCATTCGAC 477  
659 GGAACATTTATTTTACCTTAAAGACTACAGCAAAAGTTACTGAAAGTTTATGAA 718  
478 GAGGCACTCTTTTCTGTCAAAAAGCTAGAAAGAAAGTCAAGATTTCAAGTGA 537  
719 ACCCGAATGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 778  
538 ACTCAAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 597  
779 TCACCACTGTTGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 838  
598 TCTCCGCTGATCAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 657  
839 TTGCAACAAATTTGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 898  
658 ATGTACCAATTTGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 717  
899 AGTTGATGATTTGAGCTGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 958  
718 AATATATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 777  
959 CTCTGACTGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 1018  
778 TTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 837  
1019 AACTTTATCATCAGACAGAGAAATTAATTTCAAGACAACTTCCAAAGAAATTA 1078  
838 GCTCTGTCGCAAGAAAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 897  
1079 GATTATGATTTCTTCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1138  
898 GAGCTCATTTGCTTCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 957  
1139 GACCAAAATTTGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 1198  
958 TCAGTGAAGCCCAACACACCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1017  
1199 TACTACAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1258  
1018 TACTACAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1077  
1259 CAGCCAAAG 1134  
1078 CTGTTAAG--AGAGAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1194  
1319 GAGCTGCTATCTTCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1378  
1135 GAGCTGCTTCTTCAAG 1194  
1379 GACTTACGCTTCAAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1438  
1195 GCTGTGCTGAG 1254  
1439 ATTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1498  
1255 GTGACCAATCAAG 1314  
1499 TTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1371  
1315 TTTGAG 1428  
1559 GGTGAG 1618  
1372 CAAGACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1428  
1619 GCACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1678  
1429 TGCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1488  
1679 TATGAG 1738

Db 1489 GAATATGTTGCGGAGAGCTTTCTGAACTGCTTGAGAAGAGTTGACAGGTTCCATGGATT 1548  
 QY 1739 CAAATGAGAAAGCAATATGATTTGAAGTGGATGACAGAGCTGAAGCTTAAAGAGTC 1798  
 Db 1549 AATGTGAGTACCCCAAAATCATAAAGTACAGAAATCCAGCTGCATTTGTTAGAGCT 1608  
 QY 1799 TTACAGCAAAAGTTCACAGACACCCAGATAGTGTGCTGTCTGTTGTTCAAGTAATCGG 1858  
 Db 1609 ATACAGCAATATGTTGATCCTGATGTTGAGCTGGTAAATGTCATTCGCTTCTAATCAG 1668  
 QY 1859 AAGGCAAAATACGATCTATTAAAAAATACCTGTGTACAGATTCCTCCACCCCAAGTCAG 1918  
 Db 1669 AAGACCTATTATGATTCATTAATAAATATTTGAGCTCAGACTGCCAGTCCCAAGCCAA 1728  
 QY 1919 TGTGTGTCGCGGCACTTAGGCAAAACAGCAAACTGTCTATGCGCAATGCTCAAAAGATT 1978  
 Db 1729 TGTGTGTCGTCGAGCACTTGAATAAAGAGGCAATGATGATGATGATGATGATGATGATG 1788  
 QY 1979 GCCCTACAGATGAATGCAAGATGGAGGAGAGCTCTGGAGGTGACATTCCTCCCTGGAAG 2038  
 Db 1789 GCTATGCAAGTACCTGCAAGCTCGAGGCGAGCTGTGGCTGTGGAATACCTTTAAAG 1848  
 QY 2039 CTCGTGATGATGCTGTCATCGATTGTTACATGACATGACAGCTGGCGGAGGTCAATC 2098  
 Db 1849 TCCCTGATGTCGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1908  
 QY 2099 CGAGATTTGTCGACATCAATGAAGGATGACCCGCTGCTTCTCAGCTGCATATTT 2158  
 Db 1909 GTTGGATGTCGTCGAGCTTAAACCCAGCAATCAGAGTGTGTTTCCGCTGTATCCTT 1968  
 QY 2159 CAGGATAGAGGACAGAGCTGTGATGAGGCTCAAGTCTGCTGCAAGCGCTCTGAGG 2218  
 Db 1969 CAGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2028  
 QY 2219 GTTGGATAGCTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2278  
 Db 2029 AATGTGTACAGTACATCATGATGATGATGATGATGATGATGATGATGATGATGATG 2088  
 QY 2279 GGAGCGGCGAGTGAACATCTGTTGATGATGATGATGATGATGATGATGATGATGATG 2338  
 Db 2089 GGGGATGTCGTCGAGCACTTATGAATATGAAGTCCACAGCTGCTGACCACTGTG 2148  
 QY 2339 AATTCATGTTAGAGTTTACACCTAGATTAACGATTAACGATTAACGATTAACGATTA 2398  
 Db 2149 CGAGATCCAGCTCAAAATACAGCTCAAGCTGTCGGTATGTTGTCAGGAGAGTGC 2208  
 QY 2399 AACACAGATTTTGTCTCAGTCTGGAGGAGACTTCCAGATCCACTTCTCGAACAGTT 2458  
 Db 2209 ATGCCAGATTTTACCGAAATGAACCGCACTGTACAGAACCCCACTTGCACCTGT 2268  
 QY 2459 ATTGATGAGAGTTTACAGACAGATGATGATGATGATGATGATGATGATGATGATGAT 2518  
 Db 2269 GTGGATTCAGAACCAACAGCTAACGATGATGATGATGATGATGATGATGATGATGAT 2328  
 QY 2519 AGAAGTGTAGTGTCTTCCACACATTAATGTCATCTATGACAAAGCGGCTGAGG 2578  
 Db 2329 TGCGGGGAACTGTAGTCTCTACTATAATGTCATGATGATGATGATGATGATGATGAT 2388  
 QY 2579 CCAGACACATACAGCTGTGACCTACAGCTGTGACATCTATGACAAAGCGGCTGAGG 2638  
 Db 2389 CCGACCATATGACAGACTTACATTAATGTCGACCTGTGATGATGATGATGATGATGAT 2448  
 QY 2639 GTCAATGCTGCTCTGTCGTCGATGATGATGATGATGATGATGATGATGATGATGAT 2698  
 Db 2449 ATAGTGTGTCGACGACATGATGATGATGATGATGATGATGATGATGATGATGATG 2508  
 QY 2699 AGTATTCACAGAGCAAAATCTGTCTGCTCAAGTGTGCTCAAGTGTGCTCAAGTGT 2752  
 Db 2509 AGCAATTCATAAAGAACCCAGTCTGAAATAGCAACCACTCTCTTCTACCTGTGA 2562

RESULT 12

AK093475  
 LOCUS  
 DEFINITION  
 ACCSSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 FEATURES  
 source

AK093475 3138 bp mRNA linear PRI 15-JUL-2002  
 Homo sapiens cDNA FLJ36156 fis, clone TESTI2025546, weakly similar  
 to Mus musculus mRNA for MIWI (piwi).  
 AK093475  
 AK093475.1 GI:21752360  
 oligo capping; fis (full insert sequence).  
 Homo sapiens testis cDNA to mRNA, clone lib:TESTI2  
 clone:TESTI2025546.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1  
 Ishibashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S.,  
 Ono, Y., Hotuta, T., Hiraoka, S., Murakawa, K., Takiguchi, S.,  
 Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, S.,  
 Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H.,  
 Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y.,  
 Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y.,  
 Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takahashi-Fujii, A.,  
 Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,  
 Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.  
 NEDO human cDNA sequencing project  
 Unpublished  
 2 (bases 1 to 3138)  
 Isogai, T. and Yamamoto, J.  
 Direct Submission  
 Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7  
 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of  
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
 Research Association for Biotechnology (RAB); cDNA library  
 construction: Helix Research Institute (HRI) (supported by Japan  
 Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
 HRI, and Biotechnology Center, National Institute of Technology and  
 Evaluation; clone selection for full insert sequencing: HRI and  
 RAB; annotation: HRI and RAB.  
 Location/Qualifiers  
 1. 3138  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="TESTI2025546"  
 /tissue type="testis"  
 /clone lib="TESTI2"  
 /note="cloning vector: pME18SFL3"  
 152..2710  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAC04179.1"  
 /db\_xref="GI:21752361"  
 /translation="MSGRAVRKARGIARSPATVEGRIOASPLPRSDVLSNNEASSNS  
 GFLGTSRISTNDKYGISGAGDAGTSMERGVKQKQDFMDLSICTREKLAHVNRCKTSS  
 GIVKLIVTNLDFPDQDQOLYQVHTVYIPDASRRURIALLYSHLSNEAKAFDGA  
 ILFLSKLEEKVTELSSTQRTGETIKMTITRLKRLPSSPVCIQVENIIPRKILKLS  
 MYQIGNFVNPSEPMIEI FQHKLSLWPGFAISVSYFERKLLFSADVSVKVLRNTEVLEF  
 MYALCQRTGLSCTGTCTCKQLIGLIVLTRYNNRTYSIDDIIDWSVKPHTHFKEDGTEI  
 TVDYRYQOYDITVSDNLQPMVLVSLKKRDNSEAOHLIPELFCGLTQDQATSD  
 FQMKVAEKRLTSPSGQRRLRVNIOIRNTNARFELETWGLHSGLSLUTGRIVP  
 SKILMDQDHIQCPVSAADSKDITCKILNAQSLNWLILCSDRTSYVAESFLNCLR  
 VAGSGFNVDYPKIKI KOENPAAPVRAIQVYDPDQVLMCLIPSNOKTYTYSIKYL  
 SDCQVPSQCQVARTLNKQGMMSIATKAMQMTCKLGGELWAVEIPKLSLWVGDIDV  
 CKDALSKDQVWVGVCAVNPRIITWFSRCILQRTMTDADCLKVFMTGALNKWYKIDV  
 DLPARIIVYRAGVGGQLETLIEYEPQLLSVAESGNSSTSRSLSVIVVKCKMPRF  
 TEMPTVONPLGTVDSEATRNEDWDFYILISQVACRGTVSPTYNVIYDNLGLKPDH  
 MORITFKLCHLYINWPIVSPAPCOYAHKLTELVAQSIHKSPSLANHLFYL"

BASE COUNT  
 ORIGIN

912 a 678 c 714 g 834 t

Query Match  
 Best Local Similarity 56.6%;  
 Matches 1320; Conservative 0; Mismatches 1005; Indels 9; Gaps 3;

19.6%; Score 679.2; DB 9; Length 3138;  
 Pred. No. 4.8e-148;

419 GATTTCATGATCTGTTGATGATCAAGGCAAGCACTGATGTAAGTAAATCAAAA 478  
 386 GACTTTATGATTTGATGATCTGATCAAGAAATTTGCAATGTGAAATTTGAAA 445  
 479 ACAGTTCTTCAGGACTTATAGTAAAGTTAAGCACTTAACATTTCCGGCTGACATCCGT 538  
 446 ACAGTTTCCAGTGAATACCTGTGAACTGTGTAACAACTCTTTAACTTGAATTTCC 505  
 539 CCCCAGTGGCCCTTATATGATGATCACTTGAATTAACCCAGTGAAGGAGCAAGA 598  
 506 CAAGCTGGAGCTATACCAATGATGATGATGATGATGATGATGATGATGATGATG 565  
 599 CTCCTGTTAGCTCTTCTTTTCAACGAAATCTAATTTGAAGTGTGATGATGATG 658  
 566 CTGAGATGCTTCTTATATGATGATGATGATGATGATGATGATGATGATGATG 625  
 659 GGAACGATATATTTTCTTAAAGCACTAAGCAAGGATGATGATGATGATGATGATG 718  
 626 GGTGCTATCTTTTCTGATCAAAAGCTGAAAGGATGATGATGATGATGATGATG 685  
 719 ACCGGAATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 778  
 686 ACTCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 745  
 779 TCACCAATCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 838  
 746 TCTCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 805  
 839 TTGCAAAATTTGAGCAAGATTTATTTAATCCCAATGATGATGATGATGATGATG 898  
 806 ATGATCAAAATTTGAGCAAGATTTATTTAATCCCAATGATGATGATGATGATG 865  
 899 AGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 958  
 866 AATATATCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 925  
 959 CTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1018  
 926 TTTAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 985  
 1019 AACTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1078  
 986 GCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1045  
 1079 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1138  
 1046 GGGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1105  
 1139 GACGGAATCCCAAGAGCACTTTAAGAAAGCCGATCTGAGGATCACTTCTTAA 1198  
 1106 TCAATGAAAGCCCAACACACTTTTCAAGAGGAGGATGATGATGATGATGATG 1165  
 1199 TACTACAGAGAGCAATCAACAGAGATCACTGATGATGATGATGATGATGATG 1258  
 1166 TACTACAGAGAGCAATCAACAGAGATCACTGATGATGATGATGATGATGATG 1225  
 1259 CAGCCCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1318  
 1226 CTGTTAAAG--AAGAGAGAGATCAACAGAGAGGATGATGATGATGATGATG 1282  
 1319 GAGCTCTGATCTTAAAGAGATCACTGATGATGATGATGATGATGATGATGATG 1378  
 1283 GAGCTCTGATCTTAAAGAGATCACTGATGATGATGATGATGATGATGATGATG 1342  
 1379 GACTTACGCTTCTTAAAGAGATCACTGATGATGATGATGATGATGATGATGATG 1438  
 1343 GCTGTGCTGAAAG 1402  
 1439 ATTGATTAATCAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1498  
 1403 GTGGAACAATCAAG 1462  
 1499 TTTGATTCACATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 1558

1463 TTTGGAAGCCAGAT--TCTGATGATGATGATGATGATGATGATGATGATGATG 1519  
 1559 GGTGGAAG 1618  
 1520 CAAGACCAATAT--GTCAACCTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 1576  
 1619 GACACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1678  
 1577 TCAAGATTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1636  
 1679 TATGAG 1738  
 1637 GATATGTTGCGAG 1696  
 1739 CAATGAG 1798  
 1697 AATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1756  
 1799 TTAACAG 1858  
 1757 ATACAG 1816  
 1859 AAG 1918  
 1817 AAG 1876  
 1919 TGT 1978  
 1877 TGT 1936  
 1979 GGCCTTACAG 2038  
 1937 GCTATGAG 1996  
 2039 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2098  
 1997 TCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2056  
 2099 GAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2158  
 2057 GTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2116  
 2159 CAG 2218  
 2117 CAG 2176  
 2219 GCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2278  
 2177 AATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2236  
 2279 GAG 2338  
 2237 GGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2296  
 2339 AATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2398  
 2297 GCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2356  
 2399 AATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2458  
 2357 ATGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2416  
 2459 ATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2518  
 2417 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2476  
 2519 AGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2578  
 2477 TGCAG 2536  
 2579 CAG 2638







QY 1677 ATTATGAGCAGCCCAATTCATTGATACAAAATCTATTTAAAGTTACCAAGCCATGGGCA 1736  
 DB 953 CTGATATATCTCCGAGAGCTTTCGAACTGCTTGAGAAAGATTACAGGTTCCATGGGAT 1012  
 QY 1737 TGCATAATGAGAAAGCAATATGATGTAAGTGAACAGAACTGMAAGCCCTACTATTAAG 1796  
 DB 1013 TTAATGTGACTACCCCAAAATCATMAAAGTACAGAAATTCAGCTGCAATTTGTTAG 1072  
 QY 1797 TCTTACAGAAAGGTGACAGAGACACCCAGATAGTGTCTGTCTGTCTGTCAGATATC 1856  
 DB 1073 CTATACAGCAATATGTTGATCTGATGTTGATGTTGATGTTGATGTTGATGTTGATG 1132  
 QY 1857 GGAAGCAAAATACGATGCTATTTAAAAATACCTGTGACAGATTGCTTACCCCAAGTC 1916  
 DB 1133 AGAAGACCTATATGATTCATTTAAAAATATTTAGCTCAGACTGCCAGTCCCAAGCC 1192  
 QY 1917 AGTGTGTGTCGCCGACCACTTATGAGAAACACAGAACTGTCATGSCATCTCAAAA 1976  
 DB 1193 AATGTGTGTCGCTGCGACCTTGAATAAACAGGCAATGATGATGATGATGATGATGATG 1252  
 QY 1977 TTGCCCCACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2036  
 DB 1253 TCGCTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1312  
 QY 2037 AGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2096  
 DB 1313 AGTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1372  
 QY 2097 TCCGAGATTTGTTGCGAGCATCATGAAAGGATGACCCGCTGTTCTCAAGCTGATAT 2156  
 DB 1373 TTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1432  
 QY 2157 TTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2216  
 DB 1433 TTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1492  
 QY 2217 GGGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2276  
 DB 1493 ACAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1552  
 QY 2277 TAGGAGACGGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2336  
 DB 1553 TAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1612  
 QY 2337 TAAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2396  
 DB 1613 TGGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1672  
 QY 2397 TGAACACAGATTTTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 2456  
 DB 1673 GCATGCCAGATTTCTTACCGAATGAAACCGCACTGACGAACCCCACTTGGCACTG 1732  
 QY 2457 TTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2516  
 DB 1733 TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1792  
 QY 2517 TGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2576  
 DB 1793 CCTGCCGGGAACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1852  
 QY 2577 AGCCAGACCACTACAGGCTTGAACCTTCAAGCTGTGCAATCTTATTAACAATGAGC 2636  
 DB 1853 AGCCAGACCACTACAGGCTTGAACCTTCAAGCTGTGCAATCTTATTAACAATGAGC 1912  
 QY 2637 GTGTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2696  
 DB 1913 GCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1972  
 QY 2697 AGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2752  
 DB 1973 AAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2028

RESULT 14  
 AY014899 3057 bp mRNA linear INV 02-JAN-2001  
 LOCUS  
 DEFINITION Strongylocentrotus purpuratus seawater, complete cds.  
 ACCESSION  
 VERSION AY014899  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Strongylocentrotus purpuratus.  
 Strongylocentrotus purpuratus.  
 Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 Echinoidea; Echinoidea; Echinacea; Echinoida;  
 Strongylocentrotidae; Strongylocentrotus.  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 Submitted (28-NOV-2000) Biological Sciences, Rutgers  
 University-Newark, 101 Warren Street Room 135, Newark, NJ 07102,  
 USA  
 FEATURES  
 source  
 CDS  
 1..3057  
 /organism="Strongylocentrotus purpuratus"  
 /db\_xref="taxon:7668"  
 181..2745  
 /codon\_start=1  
 /product="seawater"  
 /protein\_id="AA042533.1"  
 /db\_xref="GI:12007639"  
 /translation="MDDRPGSEAPPPYQAAAPQEAASSGMPVGRGSRGTTOPPG  
 AVKQEPREGLTSQIAKAIAGSGGEGVAVDORRRRDEIELEPTREBNVYKAI  
 AGDKIALFANGFKLTKEDMDQYRDFEELINPAPRALIYVILFRRCLIM  
 DMTSLFLAKVTNLSEKSDGNNVLCVHVAATLNPAPRALIYVILFRRCLIM  
 NMEQYRRVYPTAIDIKOHGILQWGFVTSIIQYEDVLLSLDISHKLRQTVS  
 VVNDLFFNRARGFREITIKGIVLIGFVTKNNKRYRDDIDPDTPADTFETSGPVS  
 YVDYPRKGERVHDVNPMLVSRPKRBEKKGYPVILPELCLFLGLSDMDRDPG  
 VKMDGQHTRVGPQDRCRTLSGFIKULSSNEVATYIDSMEMEDKQVQLTGRVLP  
 EKLFGKQPSYNSPNDMSRDTGRNALTDKILNNKIFPTRDANKGDFIKSLVR  
 VANPMGNVIRGEIYELPDRTEYTRSLQIADQIYVILFRRCLIM  
 CVTHRPSPSOVSRVTSKQOIMSVATIKIMONCKAGDLMVEIPLNNALIKRYQIN  
 STHDSLTGRSVSLGFSVMSKSSPSSCFQHOAGSFGANLSTLNNALIKRYQIN  
 EKPEPRLIFRDGVDSQVNLVYDELKQIDTLDKYVPGSTYHKLAVVYVKKRINR  
 FPANRGGISNPPEGVIVDVVTKHLDYITIGSQSRQGSVSPSYNVVYDTTGLP  
 DHMORLYLYLTLVYNNPQVTVRVPAPCMYAKHKLAFLLIGQSVHEVPNNRLADTFYL"  
 BASE COUNT 822 a 806 c 796 g 633 t  
 ORIGIN  
 Query Match 16.3%; Score 565; DB 3; Length 3057;  
 Best Local Similarity 54.8%; Pred. No. 2.6e-121; Indels 18; Gaps 5;  
 Matches 1230; Conservative 0; Mismatches 997;  
 QY 513 CTAAACATTTCCGGCTGACATCCCGTCCCAAGTGGCTTATATCATGATCATGTTGACT 572  
 DB 512 CAACCGCTTCAAGTGAATGAACCTAAACAGACTGGAACCTACAGTACCGGGTTGACT 571  
 QY 573 ATAAACCACTGATGAAAGCCAGAAAGACTCCGTTACGCTCTTTTTCACAGAAATC 632  
 DB 572 TTGAGCCTGAGATCCTGAACCTTAGAGCTGTTTGCCTCCGAAAGGCCACAGTGC 631  
 QY 633 TAAATGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 692  
 DB 632 TTCTTGTAAAGGATTAACGCTGATGATGATGATGATGATGATGATGATGATGATG 691  
 QY 693 AAAAGTACTGAAATTTTAAAGCCCGGAATGAGATGATGAGATTAACATCA 752  
 DB 692 AAAAGTAAACCACTGAGCGCAGAGGAAAGATGGGTCTAAGTGAATATGTGTGA 751  
 QY 753 CTTTAACAAATGAATCTTCAACCTTCAACCACTTTTGAAGTTCTTATATATATTT 812  
 DB 752 CTGATGTTGATCTTAAACCCCAATGAGACCAACCTTCAACCTTCAACAGTCTCT 811

QY 813 TCAGGAGGCTTTTGAATCAATGAAATTTCCAAATAATTTGGACGAAATTTATTAACCCAA 872  
 Db 812 TTGAAGAGGTTTGAAGATGATCAATATGGAGCAAGTTTGGACGAACTTATGATCCTA 871  
 QY 873 ATGACCAATATGATATTCGAAGTTCACAGGTTGGTATTTGGCCTGGCTTCACTACTTCCA 932  
 Db 872 CGGCGCCATTCACATCAAGCAGATGGCTGACGCTGTGGCCAGGGTTGTGACATCCA 931  
 QY 933 TCTTTCAAGTATGAAACAGCATATGCTCTGACCTAGCTTTAGCCATATAAAGTCTTTCGAA 992  
 Db 932 TCTTTCAAGTATGAAACAGCATATGCTCTGACCTAGCTTTAGCCATATAAAGTCTTTCGAA 991  
 QY 993 GTGAGAGTGTCTTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 1052  
 Db 992 CCCAGACGCTGTGGGAGGTGATG---AATGATCTCTTCAACAAAGCTTGGGGGAGGTTCA 1048  
 QY 1053 AAGAAACAAGTTTCCAAAGAAATATAGGTTTATGTTTCTTACCAAGTATATAAATTAA 1112  
 Db 1049 AGGAGGAAATCACAAAGCTCATGATTGGACAGATTTGCTTCAAAAGTATCAACAAAGA 1108  
 QY 1113 CATACAGAGTGGATGATATGCTGGGACAGAAATCCCAAGAGCACCTTTTAAAGAAAGCCG 1172  
 Db 1109 CCTACCGGTTAGATGATATGACCTTTGACACTCTCCAGCGGATACCTTTTGGAGACCCGTT 1168  
 QY 1173 ACGGCTCTGGGCTCAGCTTCTTAGAATACATACAGAGCAATACCAAGAGATCACCG 1232  
 Db 1169 CGGACCA--GTTTCTATGCTGATTTCTCAAGAAAGTCTTATGAGAGGTCAITTCAG 1225  
 QY 1233 ACTTGAAGCAGCTGTCTTTGCTAGCAGCCCAAGAGAGCGGGCCCTTGGGGGAGCAC 1292  
 Db 1226 ACGTGAACCAACCCATGCTGTGTCTCGCCCAAGAGAGGGAGGAAAGAGG----- 1280  
 QY 1293 TGCCAGGSCCTGACATGCTCATCTTCTGAGCTCTGCTATCTTACAGGCTTAACTGATATAA 1352  
 Db 1281 -AGTTGCTCAGCTTACCTACCGCTCAGCTGTGCTTCTTCTGAGCTGAGCTTAAAGCAGCA 1339  
 QY 1353 TGCTTAATGATTTTAAAGTATGAAAGACTTTAGCCGTTTATCAAGAGCTTAACTCCAGAGC 1412  
 Db 1340 TGCGCGCGATTTCAACGTCATGAAGGACATGGGCCAACAACACTCTGTGTTGACCCCAAG 1399  
 QY 1413 AAAGGAGCTGAAGTGGGAGGACTCATTTGATTAATTTCAATTAATAAAGCAATATGTTCAA 1472  
 Db 1400 ACAGTGTAGAACGCTCAGCGGCTTATCAAAAAGCTCAGAGCAATGAAGAAGTGAAGA 1459  
 QY 1473 GGGAGCTTCGAGACTGGGCTTTGAGCTTTTGAATTCACAACTTACTGTCTTCCAGAGAA 1532  
 Db 1460 CTTACCTAGATTTCTGGGGCATGGAGTTTGAACAAAGCAGGTCATACTGACTGGGCGTG 1519  
 QY 1533 TTTTGAACAGAAAGATTTCAACAGGTGGAAACAACTTTGATTAACAATCCCAAAATTTG 1592  
 Db 1520 TCTTGGCCACTGAGAAACTATTTCCAGAGGGTAAACAGTTTCAATTCAGATTCGAGCAATG 1579  
 QY 1593 CAGATTGTTCCAAAGAAACAGAGGTGACCATTAATTTAGTTTGAAGCCACTAGATAACT 1652  
 Db 1580 CGACTGTTCCAGAGACACAGGGGTAATGGCCCTCAGTATCAAAAGATTCCTCAACAACT 1639  
 QY 1653 GGTGTTGATCTATACCGAAGAAATTTATGAAGCAGCAAACTTATTTGATATCAAAATCTAT 1712  
 Db 1640 GGAAGATCTTCTACACACCGCGGACGCAACAGGGGCCAGGACTTCAATCAAGTCACTCG 1699  
 QY 1713 TTAAGTTTACACCGCATGGGCTGCAATAGAGAAAGCAATTAATGATTTGAAGT---GG 1769  
 Db 1700 TAGGAGTTCGAAACCCAAATGGGTATGAACGTACGAGGACCAAGAGATAGTGGAACTTCCCG 1759  
 QY 1770 ATGACAGAACTGAAGCTTACTTTAAGAGTCTTACAGCAAAAGGTACAGCAGACACCCAGA 1829  
 Db 1760 ATGATGTTACAGAGCTTACACAGATCCCTACAGGCAAAATCGCACAGGATACAGA 1819  
 QY 1830 TAGTTGCTGCTGTTGCTGCAAGTAAATCGGAGGACAAATACGATGCTATTAAAAAATPAC 1889  
 Db 1820 TTGTTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1879

QY 1890 TGTGTACAGATTGGCTTACCCAAAGTCAAGTGTGTGTGGCCGCACTTATAGGCAACAGC 1949  
 Db 1880 GGTGTGTACCCATTCATGCCCCAGCCAGGTATCTGTCTCCAGGACCTGTCTCAAGCAGC 1939  
 QY 1950 AAATGTCTATGGCCATTTCTTACAAAGATTGCCCTTACAGATGAATCTCAAGATGGGAGGAG 2009  
 Db 1940 AGATGTCTATGTGTAGCCACCAAGATTGCCATGAGATGAATCTCAAGATGGGCGGAG 1999  
 QY 2010 AGCTCTGGAGGGTGGACATCCCCCTGAGCTCGTGTATGATCGTTGGCATCGATTGTTACC 2069  
 Db 2000 ACCTCTGGCGAGTGGAGATCCCGCTGAGCAACCTGATGATCATCGGTATTGACAGCTACC 2059  
 QY 2070 ATGACATGACAGCTGGCGGAGGTCAATCCAGATTTTGTTCAGCATCAATGAAGGA 2129  
 Db 2060 ATGACTCTTGACCAAGGCGCTCCGCTCTGGGATTCGTGGTCCATGAACAAGTCTC 2119  
 QY 2130 TGACCCGCTGGTCTTCAAGCTGATATTTAGGATAGAGGACAGGAGCTGGTAGATGGC 2189  
 Db 2120 AAACAGCTTCTTCTTAGTTGTGATTTCCAGCACACAGGAGAAATTCGGTGTCTAACC 2179  
 QY 2190 TCAAGTCTGCCGAGGCTCTGAGGCTTGGAAATAGCTGCAATGATGATACATGCCA 2249  
 Db 2180 TGAGTACCTCTGAACAATGCCCTGAAGCATATATCAGATCAACGAGAAAGTTTCTCG 2239  
 QY 2250 GCCGATCATGCTGTATCCCGCATGGCGTAGGAGACGGCCAGCTGAAACACATGGTGAAT 2309  
 Db 2240 AGCTATCATCATCTTCCGCTGATGGTGTGGTACAGCCAGTCAATCTGGTGTGACT 2299  
 QY 2310 ACGAAGTGGCCAGTTTCTGAGTTTGTCTA---AAATCCATTTGGTAGAGTTTACAACTT 2366  
 Db 2300 ATGAGCTCAAAACAATCAAGGACACCTTGGACAAAAGTCTACCTCAAGGAGCGGTTACA 2359  
 QY 2367 GANTAAAGTAAATTTGTGTGAAGAAAGAGTGAACACAGAGATTTTGTCTCAGTCTGGAG 2426  
 Db 2360 AGCTGCGGTTGT 2419  
 QY 2427 GAAGCTTTCAGATCACTTCTGGAACAGATTATGATGTAGAGTTTACAGAGCCAGAT 2486  
 Db 2420 GAGGCTGTCAACCTCCCTCCCTGGAAACCGTATCATGATGATGTGTGTGTGTGTGTGTGT 2479  
 QY 2487 GGTATGACTTTTATTTGT 2546  
 Db 2480 TCTATGACTTACTTCTATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 2539  
 QY 2547 AATGTCTATATGACAAACAGCGGCTTGAAGCCAGACCAATACAGCGCTTGAACCTACA 2606  
 Db 2540 ACAACGTTGTCTATGACACCACTGTGTCTCAAGCCAGACCAATCATCAGAGGTTGACCTACA 2599  
 QY 2607 AGCTGTGCCACATCTATTACAACTGGCCAGGTCATCTGTGTGTGTGTGTGTGTGTGTGTGT 2666  
 Db 2600 AGCTGACCATCTCTACTTCACTGTGGCGGAAACGGTGTGTGTGTGTGTGTGTGTGTGTGT 2659  
 QY 2667 AGCCCCAAGCTGGCTTTTCTTTGTGGCCAGAGTATTACAGAGAGCCAAATCTGTAC 2726  
 Db 2660 ATGCACAAAGCTGGCTTCTCTGATCGGCGAGAGCGTCCATGAAGTGTGCCCAACAGAC 2719  
 QY 2727 TGTCAAAACCGCTTTTACTACTCTA 2751  
 Db 2720 TAGCCGACACCTCTTCTATCTCTA 2744

RESULT 15  
 AX381696/c  
 LOCUS AX381696 590 bp DNA linear PAT 18-MAR-2002  
 DEFINITION Sequence 634 from Patent WO0212280.  
 ACCESSION AX381696  
 VERSION AX381696.1 GI:19576518  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1

AUTHORS Pyle, R.A., Xu, J. and Secrist, H.  
TITLE Compositions and methods for the therapy and diagnosis of colon cancer

JOURNAL Patent: WO 0212280-A 634 14-FEB-2002;  
CORIXA CORPORATION (US)

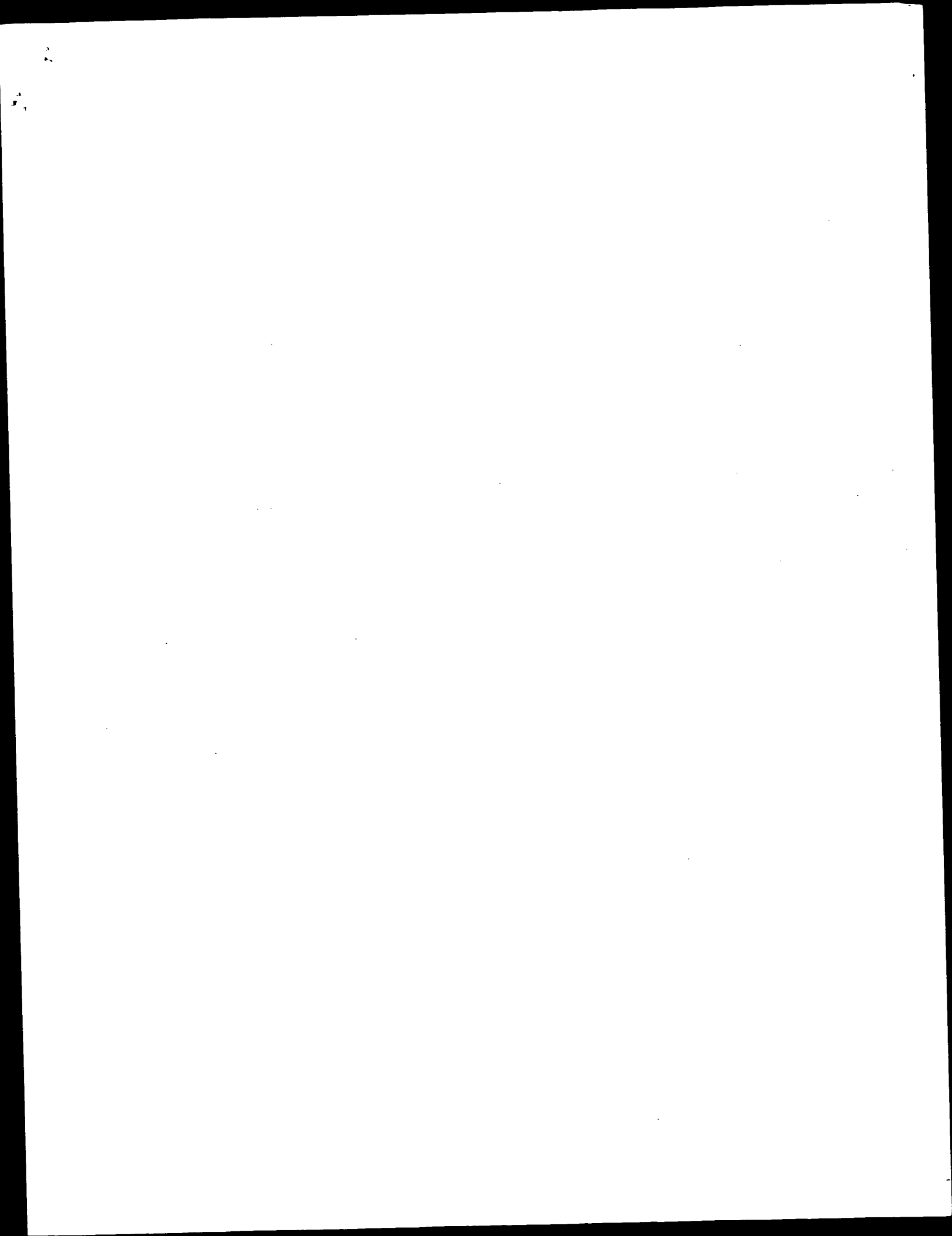
FEATURES  
source location/Qualifiers  
1. .590

BASE COUNT 209 a 105 c 65 g 208 t 3 others  
ORIGIN

Query Match 15.9%; Score 550.4; DB 6; Length 590;  
Best Local Similarity 99.3%; Pred. No. 5.7e-118;  
Matches 551; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2915 TTATTTTATATGTAATAATTAAGATTATTTATATTTATCTCTGTTGTTCTCATAGATATTT 2974  
DB 590 TTATTTTATATGTAATAATTAAGATTATTTATATTTATCTCTGTTGTTCTCATAGATATTT 531  
QY 2975 TGTGAGCATTTTGTGTTTATTTGTAAGAAATGTGATATAGATCTGTGATATTAAC 3034  
DB 530 TGTGAGCATTTTGTGTTTATTTGTAAGAAATGTGATATAGATCTGTGATATTAAC 471  
QY 3035 AGACTCTGAGAGATTTGAAATGTGTTGAGATTTACTTAAAGTACTTTCAGAGT 3094  
DB 470 AGACTCTGAGAGATTTGAAATGTGTTGAGATTTACTTAAAGTACTTTCAGAGT 411  
QY 3095 GAGCAAGTCTTATTAATTAATTAATTTATTTGAGATTAAGTACTTTCAGAGT 3154  
DB 410 GAGCAAGTCTTATTAATTAATTAATTTATTTGAGATTAAGTACTTTCAGAGT 351  
QY 3155 AAAGAGATTAAGAGCGGTAAGTAGATGCTCACTACACCATAGTGGGTTTCAGCTC 3214  
DB 350 AAAGAGATTAAGAGCGGTAAGTAGATGCTCACTACACCATAGTGGGTTTCAGCTC 291  
QY 3215 ATATCTTAAGATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3274  
DB 290 ATATCTTAAGATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 231  
QY 3275 GCTGATTTTATTTGCGAGGGGGCTAGGTTGTATGGAGTAAATAAATTAATTAAT 3334  
DB 230 GCTGATTTTATTTGCGAGGGGGCTAGGTTGTATGGAGTAAATAAATTAATTAAT 171  
QY 3335 TTTTAAATTTGTCAAAGAAACATTTTAAGCTCTTAACAAAAGGCCATGTAATTC 3394  
DB 170 TTTTAAATTTGTCAAAGAAACATTTTAAGCTCTTAACAAAAGGCCATGTAATTC 111  
QY 3395 TCTATATTAATCATCATTTATTTGTTTGAACCTGGACATGATTTCTATTTGTATA 3454  
DB 110 TCTATATTAATCATCATTTATTTGTTTGAACCTGGACATGATTTCTATTTGTATA 51  
QY 3455 AAATTAATTTGATGT 3469  
DB 50 AAATTAATTTGATGT 36

Search completed: January 9, 2003, 20:54:00  
Job time : 6591 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 18:51:34 ; Search time 463 Seconds

(without alignments)  
16887.555 Million cell updates/sec

Title: US-09-873-737A-5

Perfect score: 3472

Sequence: 1 ggagcgaaatggcgccgacgt.....taaatataatgatgcgcc 3472

Scoring table:

IDENTITY NUC  
Gapop 10.0 , Gapept 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_101002.\*  
1: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3469	99.9	3472	21	AAA07588 Human piwi gene, d
2	1833.6	52.8	4064	21	AAA07587 Mouse piwi gene, d
3	550.4	15.9	590	24	ABK55164 Human colon cancer
4	459.4	13.2	501	24	ABK44483 cDNA encoding colo
5	458.4	13.2	460	22	AA525525 Human ovarian PCR
6	395.4	11.4	423	21	AAFT1880 Human breast and o
7	377	10.9	3649	22	AA513630 cDNA sequence enco
8	329	9.5	2272	22	AAH15960 Human cDNA sequenc
9	310.2	8.9	2838	23	ABU10571 Drosophila melanog

10	286.4	8.2	300	20	AA213404	Human gene express
11	259.8	7.5	3524	23	ABU10535	Drosophila melanog
12	253	7.3	3047	21	AAA07586	Drosophila piwi ge
13	238.8	6.9	1733	22	AAH14032	Human cDNA sequenc
14	184.6	5.3	5671	23	ABU10570	Drosophila melanog
15	170.8	4.9	548	24	ABN62610	Human cancer relat
16	166.8	4.8	183	16	AAI23031	Human gene signatu
17	146.6	4.2	1026	23	AA570144	DNA encoding novel
18	137	3.9	7667	23	ABU10534	Drosophila melanog
19	111.8	3.2	2914	22	AAK53425	Human polynucleoti
20	111.8	3.2	3050	22	AAH16058	Human cDNA sequenc
21	110.6	3.2	786	22	AAH05561	Human cDNA clone (
22	101.6	2.9	325	22	AAI184676	Human polynucleoti
23	95.6	2.8	444	23	AA57327	cDNA #3 encoding p
24	91.2	2.6	601	23	AA570143	DNA encoding novel
25	91	2.5	646	23	AA577101	DNA encoding novel
26	87.2	2.5	3996	22	AAH14510	Human cDNA sequenc
27	83.8	2.4	2951	22	AAH24562	Translation initia
28	82.6	2.4	642	22	AAH07789	Human cDNA clone (
29	79.8	2.3	516	22	AAH14966	Human cDNA sequenc
30	79.6	2.3	734	20	AA215059	Human gene express
31	79.6	2.3	734	20	AA215059	Human gene express
32	79.4	2.3	771	22	AAK58246	Human immune/haema
33	78.6	2.3	213	24	AB517735	Human genome-deriv
34	70.2	2.0	1158	21	AA50398	Arabidopsis thalia
35	70.2	2.0	3195	21	AA50398	Arabidopsis thalia
36	70.2	2.0	3326	21	AA50398	Arabidopsis thalia
37	68.2	2.0	16167	24	ABU70254	Chemically treated
38	68.2	2.0	16167	24	ABU70254	Human immune syste
39	68.2	2.0	16167	24	ABU70254	Human immune syste
40	68	2.0	364	22	AB56879	Human foetal liver
41	68	2.0	364	22	AB56879	Probe #4964 for ge
42	68	2.0	364	22	AB56879	Human genome-deriv
43	64.6	1.9	7814	22	AA565149	Tumour suppressor
44	64.2	1.8	166	22	AB69444	Human foetal liver
45	64.2	1.8	166	22	AB69444	Probe #14839 for g

## ALIGNMENTS

RESULT 1	AAA07588 standard; DNA; 3472 BP.
XX	AAA07588;
XX	29-AUG-2000 (first entry)
XX	Human piwi gene, designated hiwi.
XX	Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
XX	anaemia; immunodeficiency; male infertility; human; ds.
XX	Homo sapiens.
OS	
XX	
XX	Key
XX	CDS
FT	Location/Qualifiers
FT	167..2752
FT	/tag= a
FT	/product= hiwi
FT	/transl_except= (pos:392..394; aa:Xaa)
FT	/transl_except= (pos:1073..1075; aa:Xaa)
FT	/transl_except= (pos:2369..2370; aa:Xaa)
FT	/note= "Xaa= Leu or Ile"
XX	MO200032039-A1.
XX	08-JUN-2000.
XX	03-DEC-1999; 99MO-US28764.
XX	04-DEC-1998; 98US-0110901.
XX	

PA (UYDU-) UNIV DUKE.  
 XX Lin H;  
 XX WPI; 2000-412085/35.  
 DR P-PSDB; AAY90235.  
 XX  
 PT Piwi family nucleic acids, polypeptides, and antibodies, useful in gene  
 PT therapy of diseases such as cancer and in various research and  
 PT diagnostic applications -  
 XX  
 XX Claim 19; Page 189-194; 201pp; English.  
 PS  
 CC This sequence encodes the human piwi family protein, designated  
 CC hiwi. The piwi family nucleic acids and polypeptides are used in gene  
 CC therapy of diseases such as cancer and also in various research and  
 CC diagnostic applications. The sequences can also be used to treat  
 CC tissue dystrophy, anaemia, immunodeficiency, and male infertility.  
 XX  
 SQ Sequence 3472 BP; 1044 A; 694 C; 796 G; 935 T; 3 other;

Query Match 99.9%; Score 3469; DB 21; Length 3472;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 3472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCGAATTGGGCGCCGACGTCGTCATGCTCCCGCCGCCATGGCGGCGCGGGAATTCGAT 60  
 DB 1 GGGCGAATTGGGCGCCGACGTCGTCATGCTCCCGCCGCCATGGCGGCGCGGGAATTCGAT 60

QY 61 TCCATCCTAATACGACTCACTATAGGCTCGAGCGCGCGCGGCGAGGTCTGAGGTGCA 120  
 DB 61 TCCATCCTAATACGACTCACTATAGGCTCGAGCGCGCGCGGCGAGGTCTGAGGTGCA 120

QY 121 AGAACAGGACTAGGCGGAGGCGGCGGTCGAAGAAATAGAAACAAATGACTGGGAGAGC 180  
 DB 121 AGAACAGGACTAGGCGGAGGCGGCGGTCGAAGAAATAGAAACAAATGACTGGGAGAGC 180

QY 181 CCAGCCAGAGCCAGAGGAGGCGCGGTCAGGAGACGCGCAGCTGGTGGGCTCCAC 240  
 DB 181 CCAGCCAGAGCCAGAGGAGGCGCGGTCAGGAGACGCGCAGCTGGTGGGCTCCAC 240

QY 241 TGGCAGTCAGCAACCTGGTGTATATTTCAGCTAGGCTCGAGCGGCCACACAGCAGAGGGGA 300  
 DB 241 TGGCAGTCAGCAACCTGGTGTATATTTCAGCTAGGCTCGAGCGGCCACACAGCAGAGGGGA 300

QY 301 ATTATTTGGCGGTGAGCGGAGAGAGAAACAGCAGAGAGAAACAGCCAAATGCAAGGACT 360  
 DB 301 ATTATTTGGCGGTGAGCGGAGAGAGAAACAGCAGAGAGAAACAGCCAAATGCAAGGACT 360

QY 361 CCAGATATCTGCTGGATTTTCAGAGTTATCGNTAGCAGAGAGGAGGCTCGTAGAGA 420  
 DB 361 CCAGATATCTGCTGGATTTTCAGAGTTATCGNTAGCAGAGAGGAGGCTCGTAGAGA 420

QY 421 TTTTTCATGATCTTGGTGTGAATACAGGCGAGAACCTAGACCACTGTAAAGAAATCAAAAC 480  
 DB 421 TTTTTCATGATCTTGGTGTGAATACAGGCGAGAACCTAGACCACTGTAAAGAAATCAAAAC 480

QY 481 AGGTTCTTCAGGCAATATAGTAAGGTTAAGCACTAACCAATTTCCGGCTGACATCCCGTCC 540  
 DB 481 AGGTTCTTCAGGCAATATAGTAAGGTTAAGCACTAACCAATTTCCGGCTGACATCCCGTCC 540

QY 541 CCAGTGGGCTTATATCAGTATCAGATTCAGTATACCCACTGATGGAGCCAGAGACT 600  
 DB 541 CCAGTGGGCTTATATCAGTATCAGATTCAGTATACCCACTGATGGAGCCAGAGACT 600

QY 601 CCGTTTCAGCTCTTCTTTTCAACAGCAAGATCTAATTTGAAAGTGTCTATGCTTTTGTAGG 660  
 DB 601 CCGTTTCAGCTCTTCTTTTCAACAGCAAGATCTAATTTGAAAGTGTCTATGCTTTTGTAGG 660

QY 661 AACGATATTTTACCTTAAAGACTACAGCAAGGTTTACTGAAGTTTGTAGTAAGAC 720  
 DB 661 AACGATATTTTACCTTAAAGACTACAGCAAGGTTTACTGAAGTTTGTAGTAAGAC 720

QY 721 CCGGAATGGAGGATGTGAGGATAACGATCACTTTTAAACAATGAACCTTCCACCTACATC 780  
 DB 721 CCGGAATGGAGGATGTGAGGATAACGATCACTTTTAAACAATGAACCTTCCACCTACATC 780

QY 781 ACCAACTTGTGTGAGTCTTATATATTTTTCAGGAGGCTTTTGAATAATCATGAATTT 840  
 DB 781 ACCAACTTGTGTGAGTCTTATATATTTTTCAGGAGGCTTTTGAATAATCATGAATTT 840

QY 841 GCAACAAATGGAGCAAAATTTATTAACCCAAATGACCCAAATGATATTTCAAGTCAAG 900  
 DB 841 GCAACAAATGGAGCAAAATTTATTAACCCAAATGACCCAAATGATATTTCAAGTCAAG 900

QY 901 GTTGTGATTTGGCTCGCTCTCACTACTTCCATCTTCACTATGAAACAGCATCATGCT 960  
 DB 901 GTTGTGATTTGGCTCGCTCTCACTACTTCCATCTTCACTATGAAACAGCATCATGCT 960

QY 961 CTGCACTGACCTTTAGCCATAAAGTCTTCCGAAGTGAGACTGTTTGGATTTTCATGTTCAA 1020  
 DB 961 CTGCACTGACCTTTAGCCATAAAGTCTTCCGAAGTGAGACTGTTTGGATTTTCATGTTCAA 1020

QY 1021 CTTTTCATCATCAGACAGAGAAATATTAATTTTCAAGAAACAAGTTTCCAAAGAAATAATAGG 1080  
 DB 1021 CTTTTCATCATCAGACAGAGAAATATTAATTTTCAAGAAACAAGTTTCCAAAGAAATAATAGG 1080

QY 1081 TTTAGTTTGTCTTACCAAGTATTAACATTAAGACATACAGAGTGATGATTTGACTGGGA 1140  
 DB 1081 TTTAGTTTGTCTTACCAAGTATTAACATTAAGACATACAGAGTGATGATTTGACTGGGA 1140

QY 1141 CCAGAAATCCCAAGAGCACCTTTAAGAAAGCCGCGCTCTGGGGTCAAGTCTTCTAGAAATA 1200  
 DB 1141 CCAGAAATCCCAAGAGCACCTTTAAGAAAGCCGCGCTCTGGGGTCAAGTCTTCTAGAAATA 1200

QY 1201 CTAAGGAAGCAATPACAAACAGAGATCAACGACTTGAAGACGCTGTCTTGTGACGCCA 1260  
 DB 1201 CTAAGGAAGCAATPACAAACAGAGATCAACGACTTGAAGACGCTGTCTTGTGACGCCA 1260

QY 1261 GCCCAAGAGAGGCGGCGCCCTGGGGGACACTGCCAGGCGCTGCCATCTCAATCTCTGA 1320  
 DB 1261 GCCCAAGAGAGGCGGCGCCCTGGGGGACACTGCCAGGCGCTGCCATCTCAATCTCTGA 1320

QY 1321 GCTCTGCTATCTTACAGGTCTAACTGATATAAATGCGTAATGATTTTAACTGATGAAAGA 1380  
 DB 1321 GCTCTGCTATCTTACAGGTCTAACTGATATAAATGCGTAATGATTTTAACTGATGAAAGA 1380

QY 1381 CTTAGCCGTTTCAATACAGACTTAACCTCAGAGCAAAAGCAGCGTGAAGTGGGACGACTCAT 1440  
 DB 1381 CTTAGCCGTTTCAATACAGACTTAACCTCAGAGCAAAAGCAGCGTGAAGTGGGACGACTCAT 1440

QY 1441 TGATTCATTTCAATAAAACGATAATGTTCAAGGGAGCTTCGAGACTGGGGTTTGAGCTT 1500  
 DB 1441 TGATTCATTTCAATAAAACGATAATGTTCAAGGGAGCTTCGAGACTGGGGTTTGAGCTT 1500

QY 1501 TGATTCCAACTTACTGTCTTCTCAGGAAGATTTTGCACAAACAGAAAGATTCACCAAGG 1560  
 DB 1501 TGATTCCAACTTACTGTCTTCTCAGGAAGATTTTGCACAAACAGAAAGATTCACCAAGG 1560

QY 1561 TGGAAACAACTTTGATTAATCCAAATTTGAGTTCAGATTTGGTCCAAAGAAACAGAGGTGC 1620  
 DB 1561 TGGAAACAACTTTGATTAATCCAAATTTGAGTTCAGATTTGGTCCAAAGAAACAGAGGTGC 1620

QY 1621 ACCAATTAATTTAGTGTAAAGCCACTAGATTAACCTGCTGTTGATCTATACCGGAGAAATTA 1680  
 DB 1621 ACCAATTAATTTAGTGTAAAGCCACTAGATTAACCTGCTGTTGATCTATACCGGAGAAATTA 1680

QY 1681 TGAAGCAGCCAAATTCATTTGATACAAAATCTATTTAAAGTTTACACCGCATGGCATGCA 1740  
 DB 1681 TGAAGCAGCCAAATTCATTTGATACAAAATCTATTTAAAGTTTACACCGCATGGCATGCA 1740

QY 1741 AATGAGAAAGCAATTAATGATTAAGTGGATGACAGAACTGAAGGCTTCTTAAGAGTCTT 1800  
 DB 1741 AATGAGAAAGCAATTAATGATTAAGTGGATGACAGAACTGAAGGCTTCTTAAGAGTCTT 1800

QY 1801 ACAGCAAAAGGTTCACAGCAGACACCCAGATAGTTGCTGCTGTTTCTCAAGTAATCGGAA 1860

1801 AACGAGAAAGGTCACAGCAGACACCCAGATAGTGTCTGTCTGTGTCTAAGTAACGAA 1860  
 1861 GGACAAATACGATGTCTAATAAATAATACCTGTGTACAGATTCGCCCTACCCCAAGTCACTG 1920  
 1861 GGACAAATACGATGTCTAATAAATAATACCTGTGTACAGATTCGCCCTACCCCAAGTCACTG 1920  
 1921 TGTGTGGCCCGGAACTTGAAGGCAACAGCAAACTGTCTATGCGCATTTGCTCAAAAGATTGC 1980  
 1921 TGTGTGGCCCGGAACTTGAAGGCAACAGCAAACTGTCTATGCGCATTTGCTCAAAAGATTGC 1980  
 1981 CCTACGATGAACTGTGAAGATGGGAGAGAGCTCTGAGAGGTGAGCATCCCTGAAAGCT 2040  
 1981 CCTACGATGAACTGTGAAGATGGGAGAGAGCTCTGAGAGGTGAGCATCCCTGAAAGCT 2040  
 2041 CCGATGATGATGTTGGCATGATGTTACCATGATGATGATGATGATGATGATGATGATGATG 2100  
 2041 CCGATGATGATGTTGGCATGATGTTACCATGATGATGATGATGATGATGATGATGATGATG 2100  
 2101 AGGATTTGTTGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2160  
 2101 AGGATTTGTTGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2160  
 2161 GATATGAGACAGAGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2220  
 2161 GATATGAGACAGAGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2220  
 2221 TTGGAAATAGCTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2280  
 2221 TTGGAAATAGCTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2280  
 2281 AGAGGGGACCTGAAACACCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2340  
 2281 AGAGGGGACCTGAAACACCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2340  
 2341 ATCCATTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2400  
 2341 ATCCATTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2400  
 2401 CACCAATTTTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2460  
 2401 CACCAATTTTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2460  
 2461 TGATGTAGAGGTTCACAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2520  
 2461 TGATGTAGAGGTTCACAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2520  
 2521 AAGTGTAGTGTCTCTCCACACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2580  
 2521 AAGTGTAGTGTCTCTCCACACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2580  
 2581 AGACCAATTAACAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2640  
 2581 AGACCAATTAACAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2640  
 2581 AGACCAATTAACAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2640  
 2641 CATTGTGTCTGTCTCTCCACACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2700  
 2641 CATTGTGTCTGTCTCTCCACACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2700  
 2701 TATTACAGAGAGCAAACTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2760  
 2701 TATTACAGAGAGCAAACTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2760  
 2701 TATTACAGAGAGCAAACTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2760  
 2761 AGACGATGAGAGCTGTCTCTCCACACATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 2820  
 2761 AGACGATGAGAGCTGTCTCTCCACACATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 2820  
 2821 CTTTTTTTTTAATCTGTATCTTTTGTGATGAACTTTGGAGAGGGATTAAGGATCTAGC 2880  
 2821 CTTTTTTTTTAATCTGTATCTTTTGTGATGAACTTTGGAGAGGGATTAAGGATCTAGC 2880  
 2881 ATTTTATTTCTAGCATGTGATTAACCGGCTCTTATTTTATGTAATAATTAAGATT 2940  
 2881 ATTTTATTTCTAGCATGTGATTAACCGGCTCTTATTTTATGTAATAATTAAGATT 2940

2881 ATTTTATTTCTAGCATGTGATTAACCGGCTCTTATTTTATGTAATAATTAAGATT 2940  
 2941 TTTATTTTATTTCTGTTCTGTTCTCATAGATATTTGAGCATTTTGTATTTTGA 3000  
 2941 TTTATTTTATTTCTGTTCTGTTCTCATAGATATTTGAGCATTTTGTATTTTGA 3000  
 3001 AGAAATGTGATTAAGATTAAGTGTGATTAATAAACAATCTCTGAGAGATTTTGAATGT 3060  
 3001 AGAAATGTGATTAAGATTAAGTGTGATTAATAAACAATCTCTGAGAGATTTTGAATGT 3060  
 3061 GTTTGAGATTTTACTTAACGATCTTTCAGAGTGAAGCAAGTCTCTTAACCTATA 3120  
 3061 GTTTGAGATTTTACTTAACGATCTTTCAGAGTGAAGCAAGTCTCTTAACCTATA 3120  
 3121 TTAATTTATTTTGAATTAAGTGTGATTAATAAAGATTAAGAGCGTAAAGTGG 3180  
 3121 TTAATTTATTTTGAATTAAGTGTGATTAATAAAGATTAAGAGCGTAAAGTGG 3180  
 3181 ATGCTCACTACACATTAAGTGTGATTAAGTGTGATTAAGTGTGATTAAGTGTGATTAAGTGTGAT 3240  
 3181 ATGCTCACTACACATTAAGTGTGATTAAGTGTGATTAAGTGTGATTAAGTGTGATTAAGTGTGAT 3240  
 3241 ATATTAATTAATTAACAGATTAAGTGTGATTAAGTGTGATTAAGTGTGATTAAGTGTGATTAAGTGT 3300  
 3241 ATATTAATTAATTAACAGATTAAGTGTGATTAAGTGTGATTAAGTGTGATTAAGTGTGATTAAGTGT 3300  
 3301 TAGTGTGATGAGATTAATAAATTTTAAATTTTAAATTTTCCAAAGAAATTTT 3360  
 3301 TAGTGTGATGAGATTAATAAATTTTAAATTTTAAATTTTCCAAAGAAATTTT 3360  
 3361 AAGATCTTTTAAACAAAGGCAATGATTAATCTATATTAATTAATTAATTAATTAATTTT 3420  
 3361 AAGATCTTTTAAACAAAGGCAATGATTAATCTATATTAATTAATTAATTAATTAATTTT 3420  
 3421 GTTTGAGATGAGATTAATTTTGTATTAATAATTAATTAATTAATTTT 3472  
 3421 GTTTGAGATGAGATTAATTTTGTATTAATAATTAATTAATTAATTTT 3472

RESULT 2  
 AAA07587  
 ID AAA07587 standard; DNA; 4064 BP.  
 AC AAA07587;  
 XX  
 XX  
 DT 29-AUG-2000 (first entry)  
 XX  
 XX  
 DE Mouse piwi gene, designated miwi.  
 XX  
 XX  
 KW Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;  
 KW anaemia; immunodeficiency; male infertility; mouse; ds.  
 OS Mus sp.  
 XX  
 XX  
 FH Key location/Qualifiers  
 CDS 191..2779  
 FT /tag= a  
 FT /product= miwi  
 FT /transl\_except= (pos:450..452; aa:Xaa)  
 FT /transl\_except= (pos:1337..1339; aa:Xaa)  
 FT /transl\_except= (pos:2636..2638; aa:Xaa)  
 FT /note= "Xaa= Leu or Ile"  
 FT /transl\_except= (pos:836..838; aa:Xaa)  
 FT /note= "Xaa= unspecified amino acid"  
 XX  
 PD MO200032039-A1.  
 XX  
 XX  
 PD 08-JUN-2000.  
 XX  
 XX  
 PF 03-DEC-1999; 99WO-US28764.  
 XX  
 XX  
 PR 04-DEC-1998; 98US-0110901.  
 XX

PA (UYDU-) UNIV DUKE.

XX Lin H;

DR WPI; 2000-412085/35.

DR P-PSDB; AAY90234.

XX Piwi family nucleic acids, polypeptides, and antibodies, useful in gene therapy of diseases such as cancer and in various research and diagnostic applications -

XX Claim 19; Page 180-185; 201pp; English.

XX This sequence encodes the mouse piwi family protein, designated miwi. The piwi family nucleic acids and polypeptides are used in gene therapy of diseases such as cancer and also in various research and diagnostic applications. The sequences can also be used to treat tissue dystrophy, anaemia, immunodeficiency, and male infertility.

XX Sequence 4064 BP; 1114 A; 978 C; 1077 G; 890 T; 5 other;

Query Match 52.8%; Score 1833.6; DB 21; Length 4064;

Best Local Similarity 81.8%; Pred. No. 0;

Matches 2125; Conservative 0; Mismatches 471; Indels 3; Gaps 1;

QY 164 ACAATGACTGGAGAGCCGAGCCAGAGCCAGAGAGGCGCGGTTCAGGACAGCG 223

DB 188 AATATGACTGCCAGGCCGAGCTCGGGCCGCGCGAGGACAGAGTTCAGGAGCGGTG 247

QY 224 CAGCTGGTGGCTCCACTGCGCAGTTCAGCAACCTGGTATATTCAGGCTAGGCTCAGCGG 283

DB 248 CAGCATGTGGGGCTGCTCGCAGCCAGCAACCTGGGTACATCCACCGAGACCTCAACAG 307

QY 284 CCACACAGAGGGGGAATTTATTTGGCCGTGGAGCGGACAGAGAGGAA---CAGCAGAGGA 340

DB 308 TCCCCCAGAGGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGTCGGAGCC 367

QY 341 ACAGCCAGTCAAGGACTCCAGATATCTGCTGGATTTTCAGGAGTTATCGNTAGCAGAG 400

DB 368 ACATCCAAAGTCACAAAGTTCAGATCTCAGCTGGTTCAGGAGCTGCTCAGGACAGAG 427

QY 401 AGAGAGGTGCTGTAGAGATTTTCATGATCTTGGTGTGAATCAAGGACAGACCTAGAC 460

DB 428 AGAGAGGGGTGCGCGAGACTTCCATGACNTTGGTGTGAACACACAGACAGAACTTGCAC 487

QY 461 CATGTTAAAGATCAAAAACAGGTTCTTCAGGCAATATAGTAAGTTAAGCACTAACCAT 520

DB 488 CATGTCAAAGAGTCAAGACAGGCTCCTCTGGCATCATTTGGAAGCTGAGCACCACAC 547

QY 521 TTCGGCTGACATCCCGTCCCGAGTGGGCTTTATATCAGTATCAGTATCAGTATTAACCCA 580

DB 548 TTCGGCTGACCTCGCGCCACAGTGGGCCCTGTATCAGTATCAGTATTAACCCA 607

QY 581 CTGATGAGAGCAGAGATCTCGGTCAGCTCTCTTTTTCACACAGAGATCTAATGGA 640

DB 608 CTGATGAGAGCAGAGATCTCGGTCAGCTCTCTTTTTCACACAGAGATCTAATGGA 667

QY 641 AAGTGTGATCTTTGATGGAACGATTTATTTTACCTTAAAGACTACAGCAAGGTT 700

DB 668 AGTGTGATCTTTGATGGAACGATTTATTTTACCTTAAAGACTACAGCAAGGTT 727

QY 701 ACTGAAGTTTGTAGTAAGACCCGGAATGAGAGGATGTGAGGATAACGATCACTTTAACA 760

DB 728 ACAGAAGTATTCAGTCAGACTCGGAATGGGAAACAGTGAGGATCACTCACTGACC 787

QY 761 AATGAATCTCCACCTACATCACCCTGTTTGGAGTTCTATATATTTATTTTTCAGAGG 820

DB 788 AACGAGCTGCGGCCACCTCGGCCACCTGCGGAGTTCTATATATCTTTCAGAGG 847

QY 821 CTTTTGAAATCATGAATTTGCAACAAATTTGACGAAATTTATTAACCCAAATGACCCA 880

DB 848 CTTCTGAAATCATGAATTTGCAACAAATTTGACGAAATTTATTAACCCAAATGACCCC 907

QY 881 ATTGATATTCAGTCAAGTTCAGGTTGGTGGCTTGGCTTCACTACTTCCATCTTCAG 940

DB 908 ATTGATATTCAGTCAAGTTCAGGTTGGTGGCTTGGCTTCACTACTTCCATCTTCAG 967

QY 941 TATGAAACAGCATCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1000

DB 968 TATGAAACAGCATCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1027

QY 1001 GTTTTGGATTTCACTTTTATCATCATCATCATCATCATCATCATCATCATCATCATCAT 1060

DB 1028 GTCTTACATTTTCACTTTTATCATCATCATCATCATCATCATCATCATCATCATCAT 1087

QY 1061 GTTTTCAAGAAATTAATGAGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 1120

DB 1088 GTGTGGAAGAGGCTCATGAGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTT 1147

QY 1121 GTGATGATATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1180

DB 1148 GTGATGATATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1207

QY 1181 GGGTCAAGCTTTTGTGATATCTACAGAAAGCAATCAACCAAGAGATCACCGACTTGAAG 1240

DB 1208 GAGTCAAGCTTTTGTGATATCTACAGAAAGCAATCAACCAAGAGATCACCGACTTGAAG 1267

QY 1241 CAGCTGTCTTGTGCTAGCGCCCAAGAGAGCGGGGCTTGGGGGACACTTGCAGGG 1300

DB 1268 CAGCGGTGTGTGAGCCCAAGAGAGCGGGGCTTGGGGGACACTTGCAGGG 1327

QY 1301 CTTGCCATGCTCATCTCAGCTCTGCTTCTTACAGTCTTAACTGATTAATAATGCTTAAT 1360

DB 1328 CAGCTATGCTCATCTCAGCTCTGCTTCTTCAAGCTTCTGCTTCTGCTTCTGCTTCTGCT 1387

QY 1361 GATTTTACGCTGATGAAGACTTAGCCGTTTCAATACAGACTTATCCAGAGCAAGGCGAG 1420

DB 1388 GATTTTCAATGTGATGAAGGACCTGCGAGTGACACCGCGCTGACCCCTGAGCAGCGGAG 1447

QY 1421 CGTGAAGTGGAGGACTCATTTGATTTTCAATTAATAAGATTAATGTTTCAAGGAGGCTT 1480

DB 1448 CGGAGGTGGGGGCTTCACTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCT 1507

QY 1481 CGAGACTGGGGTTCAGCTTTGATTTTCAACTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCT 1540

DB 1508 CGAGACTGGGGTTCAGCTTTGATTTTCAACTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCT 1567

QY 1541 ACAGAAAGATTCACCAAGTGGGAAAAAATTTGATTAATAATTCACAAATTTGCAAGTTGG 1600

DB 1568 TCTGAGAGATCCAGCGGGGGAAGACGTTTGTATTACAACTTGTGAGAGTTGGAGCTGG 1627

QY 1601 TCCAAAGAAACAGAGGTCACCATTAATTTAGTGTGTTAGCCACTAGATACTGGCTGTTG 1660

DB 1628 TCCAAAGAAACAGAGGTCACCATTAATTTAGTGTGTTAGCCACTAGATACTGGCTGTTG 1687

QY 1661 ATCTATACGCGAAGAAATTTATGAAGCAGCCAAATTTTGTATACAAATCTATTAAAGTT 1720

DB 1688 ATCTATACGCGAAGAAATTTATGAAGCAGCCAAATTTGTATACAACTTGTGAGAGTTG 1747

QY 1721 ACACAGGCTGCGGATCCAAATGAGAAAGCAATTAATTTAGTGTGAGTTGAGTGAAGTGG 1780

DB 1748 ACTCAGGCTTGGGATCCAGATGAAAGGCAATCATGATCGAGGTGGATGACAGAACTA 1807

QY 1781 GAGGCTTACTTAAAGTCTTACAGCAAAAGGTCACAGCAGCAGCAGCAGCAGCAGCAGCAG 1840

DB 1808 GAGGCTTACTTAAAGTCTTACAGCAAAAGGTCACAGCAGCAGCAGCAGCAGCAGCAGCAG 1867

QY 1841 GTTGTGTCAAGTAAATTCGGAAGGCAAAATACGATGCTATTAAAAAATACCTGTGTACAGAT 1900

DB 1868 GTTGTGTCAAGTAAATTCGGAAGGCAAAATACGATGCTATTAAAAAATACCTGTGTACAGAT 1927

QY 1901 TGCCCTTACCCCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1960

DB 1928 TGCCCTTACCCCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1987

QY 1961 GCCATTGCTACAAAGATTGCCCTTACAGATGAACTGCAAGATGGGAGGAGGCTCTGGAGG 2020



```

Db 1988 GCCATTGCCACCAAGATCGCCCTGACATGAACTGCAATGGAGCGAGCTCTGGCG 2047
Qy 2021 GTGACATCCCCCTGAAAGCTCGATGATCGTTGGCATGATTTTACATGACATGACA 2080
Db 2048 GTGACATGGCCCTGAAACGCGCATGATCGTGGCATGCACTGTTACATGACACCA 2107
Qy 2081 GCTGGGGGAGAGTCATCCGAGATTTTGTCCAGATCAATGAAGGATGACCCGCTGG 2140
Db 2108 GCTGGGGGAGAGTCATCCGAGATTTGTCCAGATCAATGAAGGATGACCCGCTGG 2167
Qy 2141 TTCTCAAGCGCATATTTTACAGATAGAGACAGAGACTGGTATGAGTGGCTCAAGCTGC 2200
Db 2168 TTCTCCGCTGCTCTTTACAGACCCGCGGACAGAGACTGGTATGATGCTCAAGGTGTC 2227
Qy 2201 CTGCAAGCGGCTCTGAGGGCTTGAATAGCTGATGATGATGATGATGATGATGATGATG 2260
Db 2228 TTGCAAGCTGCTCTGAGGGCTTGAATAGCTGATGATGATGATGATGATGATGATGATG 2287
Qy 2261 GTGATCCGCAATGGCGTGAAGACGCGCAAGCTGATAAACACTGGTGAACCTGAGAGTCCA 2320
Db 2288 GTGATCCGCAATGGCGTGAAGACGCGCAAGCTGATAAACACTGGTGAACCTGAGAGTCCA 2347
Qy 2321 CAGTTTGGATGCTTAATATCCATTTGATGATGATGATGATGATGATGATGATGATGATG 2380
Db 2348 CAGTTTGGATGCTTAATATCCATTTGATGATGATGATGATGATGATGATGATGATGATG 2407
Qy 2381 GTGATGAAGAAAGATGAACACCAATTTTGTGCTGATGCTGAGAGAGATTTTCAAGAT 2440
Db 2408 GTGATGAAGAAAGATGAACACCAATTTTGTGCTGATGCTGAGAGAGATTTTCAAGAT 2467
Qy 2441 CCATCTCTGGAACAGTTATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2500
Db 2468 CCTCTTCCAGGACAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2527
Qy 2501 ATCTGAGCCAGGCTGTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2560
Db 2528 ATCTGAGCCAGGCTGTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2587
Qy 2561 GACAAACAGCGCTGAAAGCCAGACCAATACAGGCTTGAACCTGACAGCTGACAGCATC 2620
Db 2588 GACAAACAGCGCTGAAAGCCAGACCAATACAGGCTTGAACCTGACAGCTGACAGCATC 2647
Qy 2621 TATTACAACTGCGCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2680
Db 2648 TATTACAACTGCGCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2707
Qy 2681 GCTTTTCTTTGTGGCCAGAGATTTCAAGAGAGCCAAATCTGTCACTGTCAAAACCGCTT 2740
Db 2708 GCTTTTCTTTGTGGCCAGAGATTTCAAGAGAGCCAAATCTGTCTCCCTGCAACCGCTT 2767
Qy 2741 TACTAACCTTAACCTGCG 2759
Db 2768 TACTAACCTTAACCAAG 2786

```

RESULT 3  
ID ABRK5164 standard; cDNA; 590 BP.

ABR5164;

18-JUN-2002 (first entry)

Human colon cancer-associated cDNA, SEQ ID No 634.

Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.

Homo sapiens.

WO200212280-A2.

14-FEB-2002.

```

XX 30-JUL-2001; 2001WO-US23826.
XX
XX PR 03-AUG-2000; 2000US-223265P.
XX PR 02-OCT-2000; 2000US-227406P.
XX PR 20-MAR-2001; 2001US-277459P.
XX PR 03-JUL-2001; 2001US-302702P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Pyle RA, Xu J, Secretist H;
XX WPI; 2002-257462/30.
XX
XX Novel polynucleotide encoding colon tumour polypeptides, useful as
XX vaccines for treating colon cancers
XX
XX Claim 1; Page 329-330; 425pp; English.
XX
XX The invention relates to isolated polynucleotides (i) encoding colon
XX tumour polypeptides (ii). (i) is useful for stimulating an immune
XX response in a patient and treating colon cancer in a patient.
XX Oligonucleotides derived from (i) are useful for determining the presence
XX of cancer in a patient. (i) and (ii) are useful in pharmaceutical
XX compositions, e.g. vaccines, and other compositions for the diagnosis
XX and treatment of colon cancer. A composition comprising a first component
XX selected from physiologically acceptable carriers and immunostimulants,
XX and an antigen-presenting cell expressing (ii) is useful for inhibiting
XX development of cancer in a patient. (i) is useful in the design and
XX preparation of ribozyme molecules for inhibiting expression of tumour
XX polypeptides and (i). ABRK4531-ABR5464 represent human colon cancer cDNA
XX sequences of the invention.
XX
XX Sequence 590 BP; 209 A; 105 C; 65 G; 208 T; 3 other;
XX
XX Query Match 15.9%; Score 550.4; DB 24; Length 590;
XX Best Local Similarity 99.3%; Pred. No. 9.2e-128;
XX Matches 551; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
Qy 2915 TATTATTATGTAATAAATTAAGATTTTATATTTTATCTTGTGTCATAGATATT 2974
Db 590 TATTATTATGTAATAAATTAAGATTTTATATTTTATCTTGTGTCATAGATATT 531
Qy 2975 TGTGAGCATTTTGTGTTTATTTTGAAGAAATGTGGATTAACCTTGGTATTAAC 3034
Db 530 TGTGAGCATTTTGTGTTTATTTTGAAGAAATGTGGATTAACCTTGGTATTAAC 471
Qy 3035 AGACTCTCGAGAGTATTGAAGATGTTTGAAGATTTACTTAACCTTAAGAGT 3094
Db 470 AGACTCTCGAGAGTATTGAAGATGTTTGAAGATTTACTTAACCTTAAGAGT 411
Qy 3095 GAGCAAGTCTTAATAAACCCTATATTACTTTATTTTGAATACCTGTTTGAATTT 3154
Db 410 GAGCAAGTCTTAATAAACCCTATATTACTTTATTTTGAATACCTGTTTGAATTT 351
Qy 3155 AAGAGATTAAGAGCGTAAATAGAGCTCCTACACCAACCTAGGGGTTTCAAGCTC 3214
Db 350 AAGAGATTAAGAGCGTAAATAGAGCTCCTACACCAACCTAGGGGTTTCAAGCTC 291
Qy 3215 ATATCTTAAGATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3274
Db 290 ATATCTTAAGATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 231
Qy 3275 GCTTGAATTTTATTTTGGCAGGGGGCTAGGTTGATGGAGTAAATAAATTTGAAAT 3334
Db 230 GCTTGAATTTTATTTTGGCAGGGGGCTAGGTTGATGGAGTAAATAAATTTGAAAT 171
Qy 3335 TTTTAAATGTCGAAAGAAACATTTTAAAGCTCTTAAACAAATAAGCCATAGTAATC 3394
Db 170 TTTTAAATGTCGAAAGAAACATTTTAAAGCTCTTAAACAAATAAGCCATAGTAATC 111
Qy 3395 TCTATATTAAACATCACTATTATTGTTTGAACCTGGACATGATGATTTGTTATA 3454

```

Db 110 TCTATATTAACATTACTATTATTTTGTGGAACCTGGACATGATCTATTGTTATA 51  
 QY 3455 AAATAAAATTGATGT 3469  
 Db 50 AAATAAAATTGATGT 36

## RESULT 4

ABK44483  
 ID ABK44483 standard; cDNA; 501 BP.  
 XX  
 AC ABK44483;

XX  
 DT 05-JUN-2002 (first entry)

XX cDNA encoding colon tumour protein, SEQ ID No 34.

XX Human; colon tumour; vaccine; colon cancer; immunogenic;  
 KW immunotherapy; gene; ss.

XX Homo sapiens.

XX WO200212328-A2.

XX 14-FEB-2002.

XX 31-JUL-2001; 2001WO-US24218.

XX 03-AUG-2000; 2000US-23283P.

XX 28-MAR-2001; 2001US-279763P.

XX 29-JUN-2001; 2001US-302051P.

XX (CORI-) CORIXA CORP.

XX King GE, Meagher MJ, Xu J, Secrist H;

XX WPI; 2002-241739/29.

XX New colon cancer polypeptides and polynucleotides, useful as vaccines,  
 PT for diagnosing, preventing, and treating colon cancer, and as markers  
 PT for the progression of cancer -

XX Claim 1; SEQ ID No 34; 147pp; English.

XX The invention relates to polynucleotides encoding colon tumour proteins.  
 CC The polynucleotides and encoded polypeptides are useful in pharmaceutical  
 CC compositions, such as vaccines, for the diagnosis, prevention, and  
 CC treatment of colon cancer. Polynucleotide sequences may be used as  
 CC hybridisation probes or primers, and in the design and preparation of  
 CC ribozyme molecules for inhibiting expression of tumour polypeptides and  
 CC proteins in tumour cells. The compositions are useful for stimulating an  
 CC immune response against cancer, particularly for the immunotherapy of  
 CC colon cancer, and as markers for the progression of cancer.  
 CC ABK44450-ABK46237 represent coding sequences of human colon tumour  
 CC proteins of the invention.

CC Note: With the exception of SEQ ID No 1 and 2, the sequence data  
 CC for this patent did not form part of the printed specification but was  
 CC supplied by the European Patent Office.

XX Sequence 501 BP; 134 A; 81 C; 93 G; 193 T; 0 other;

Query Match 13.2%; Score 459.4; DB 24; Length 501;  
 Best Local Similarity 99.8%; Pred. No. 5.6e-105;  
 Matches 460; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2695 CCAGAGTATTCACAGAGGCAATCTGTCAGTCAACCGCCTTTACTACTTAACC 2754

Db 1 CCAGAGTATTCACAGAGGCAATCTGTCAGTCAACCGCCTTTACTACTTAACC 60

QY 2755 TGCAGAGAGGATGACCGCGCTTTCTTTTGAATGACTTTGGGATTTTAAAGCTTT 2814

Db 61 TGCAGAGAGGATGACCGCGCTTTCTTTTGAATGACTTTGGGATTTTAAAGCTTT 120

QY 2815 TATTTACTTTTTTAACTGTTATCTTCTCGATGAACCTTGGGAAGGGATTAGGAGA 2874  
 Db 121 TATTTACTTTTTTAACTGTTATCTTCTCGATGAACCTTGGGAAGGGATTAGGAGA 180  
 QY 2875 TCTAGCATTTTATTTCTAGCATTTGCTATTTCACCGGCTTCCCTTATTATATGTAATAAATT 2934  
 Db 181 TCTAGCATTTTATTTCTAGCATTTGCTATTTCACCGGCTTCCCTTATTATATGTAATAAATT 240  
 QY 2935 AAGATTTTATATTTTATCTTCTTCTTCTCATAGATATTTTGTGAGCATTTTGTGTTA 2994  
 Db 241 AAGATTTTATATTTTATCTTCTTCTTCTCATAGATATTTTGTGAGCATTTTGTGTTA 300  
 QY 2995 TTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTG 3054  
 Db 301 TTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTG 360  
 QY 3055 AAATGTGTTTGGAGATTTTACTTAAACGTTACTTTCAGGAGTGAGCAAGTCCCTACTTATAAA 3114  
 Db 361 AAATGTGTTTGGAGATTTTACTTAAACGTTACTTTCAGGAGTGAGCAAGTCCCTACTTATAAA 420  
 QY 3115 CCTATATTAACCTTTATTTTGTGAGATACCTGTTTGAATTTA 3155  
 Db 421 CCTATATTAACCTTTATTTTGTGAGATACCTGTTTGAATTTA 461

## RESULT 5

AAS25525/c

ID AAS25525 standard; cDNA; 460 BP.

XX AAS25525;

XX 07-NOV-2001 (first entry)

XX Human ovarian PCR-subtracted cDNA library clone #1610.

XX Immunogenic protein; cancer; ovarian tumour; T-cell stimulation; ss;

XX gene therapy; cytostatic; T-cell expansion; nucleic acid hybridisation;

XX primer; probe.

XX Homo sapiens.

XX WO200157207-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US03733.

XX 04-FEB-2000; 2000US-0180403.

XX 28-MAR-2000; 2000US-0192745.

XX (CORI-) CORIXA CORP.

XX Algate PA, Mannion J;

XX WPI; 2001-488875/53.

XX New polynucleotides encoding ovarian tumour proteins, useful for

XX treating ovarian cancer, and as probes, primers, and markers of cancer

XX progression -

XX Example 1; page 374; 378pp; English.

XX The invention comprises compositions used for the therapy and diagnosis  
 CC of ovarian cancer. The compositions comprise one or more ovarian tumour  
 CC proteins, their associated polynucleotides, or immunogenic portions of  
 CC the proteins. The ovarian tumour polynucleotides and polypeptides are  
 CC useful for stimulating and/or expanding T cells specific for a tumour  
 CC protein. They are also useful for inhibiting the development of cancer in  
 CC a patient with an ovarian tumour DNA or protein by incubating isolated  
 CC T-cells allowing them to proliferate, and administering to the patient.  
 CC The sequences can be used as markers for cancer, for example, to monitor  
 CC ovarian cancer progression. Probes and primers are useful in nucleic acid  
 CC hybridisation, in detecting the presence of complementary sequences in a



XX 18-DEC-2001 (first entry)  
 DT cDNA sequence encoding mammalian Spg16.  
 XX Mammalian; reproductive-specific protein; male infertility; gene therapy;  
 KW spermatogenesis; sperm count disorder; anti infertility;  
 KW reproduction; ss.  
 XX Mammalia.  
 OS WO200166752-A2.  
 PN 13-SEP-2001.  
 PD 07-MAR-2001; 2001WO-US07371.  
 XX 07-MAR-2000; 2000US-0187518.  
 PR 12-JAN-2001; 2001US-0261557.  
 XX (WHED ) WHITEHEAD INST BIOMEDICAL RES.  
 PA Wang PJ, Page DC;  
 XX WPI: 2001-570774/64.  
 DR P-PSDB; AAU07866.  
 DR Novel reproduction-specific protein, useful for treating disorders of  
 PT reduced sperm count, enhancing/increasing sperm count and/or sperm  
 PT activity -  
 XX Claim 1; Fig 15; 151pp; English.  
 PS The present invention relates to the isolation of novel mammalian and  
 CC human reproductive-specific proteins (AAU07859-AAU07899), and the  
 CC nucleic acids encoding them. The nucleic acids encoding  
 CC reproductive-specific proteins are useful for diagnosing infertility  
 CC which is a result of reduced sperm count, reduced sperm motility,  
 CC malformed sperm or combinations of these. The sequences of the invention  
 CC are useful as markers for spermatogonial cells, for identifying genes or  
 CC proteins characteristic of male infertility, diagnosing or aiding in  
 CC the diagnosis of infertility in men, and for contraception in which  
 CC sperm production or sperm count is reduced or defective sperm is  
 CC produced. Antibodies to reproductive-specific proteins are useful for  
 CC determining the presence of these proteins in a sample obtained from a  
 CC man being assessed for infertility, for identifying the expression of  
 CC genes in particular cell type or particular developmental stage, for  
 CC studies of spermatogenesis, and for immunofluorescence of germ cells or  
 CC in Western blots for assessing the presence of the protein the antibody  
 CC binds. The sequences of the invention are also useful for treating  
 CC disorders of reduced sperm count, and for increasing sperm count and/or  
 CC sperm activity. The nucleic acids of the invention are useful in gene  
 CC therapy. AAS13623-AAS13647 represent cDNA sequences encoding for the  
 CC mammalian reproduction-specific proteins of the present invention.

XX Sequence 3649 BP; 947 A; 841 C; 911 G; 950 T; 0 other;  
 SQ Query Match 10.9%; Score 377; DB 22; Length 3649;  
 Best Local Similarity 52.5%; Pred. No. 7.1e-84;  
 Matches 974; Conservative 0; Mismatches 862; Indels 21; Gaps 6;

QY 914 CTGGCTTCACTTCCATCTTCCATGATGAAACAGCATCATGCTGCACTACGCTT 973  
 DB 1 CTGGCTATGCGCGTAGTATCCCGAGGACAGACGGGGGCTCTTCTGCTGCTGATGTC 60  
 QY 974 AGCCATAAAGTCTTCCAAAGTGAGACTGTTTGGATTTTCATGTTCAACTTTTATCATCAG 1033  
 DB 61 TCTCATTAAGTCAATTCGGACGACTCTGTGCTGGATGTCATGCTATCTACACAG 120  
 QY 1034 ACAGAAGACATAAATTTTCAAGAACAAAGTTTCCAAAGAAATTAATAGTTTGTCTT 1093  
 DB 121 AACAGGAGCAC---TTCCAGGACGATGTCAGCAAGCTTCTGTTGGCAGCATTTGTCATC 177

QY 1094 ACCAAGTATAACATTAAGACATACAGAGTGGATGATATTGCTGGGACCAAGATCCCAAG 1153  
 DB 178 ACGGCTACAAACATCGTACTACCGAATCGATGATGGAGTGGAAACAAGACCCCTAAA 237  
 QY 1154 AGCACCTTTAAGAAAGCGCGCTCTGGGGTCTAGCTTCTTAGAATACTACAGGAAGCAA 1213  
 DB 238 GACAGCTTTGTGTCATGTCGGAGCGGAAGAAATCACATTCCTGGAATACTACAGCAAAAC 297  
 QY 1214 TACAACCAAGAGATCACCGACTTGAAGCAGCTGTCTTGGTCAAGCCAGCCCA---AGAGA 1270  
 DB 298 TATGGGATCAAGTCAAGGAAGATGACACCGCTGTGATCCACCGGCCAGTGAGAGA 357  
 QY 1271 AGCGGGGCTGGGGGACACTGCCAGGCGCTGCCATGCTCATTCCTGAGCTCTGCTAT 1330  
 DB 358 CAGAAATACATGGCATGTTGCTGAAGGGGAGATCTCTGCTGCTGCGCGAGCTCTCTTC 417  
 QY 1331 CTTACAGGTCTAACTGATAAAATCGGTAATGATTTTAACTGATGAAGACTTTAGCGGTT 1390  
 DB 418 ATGACGGGATCCCTGAGAAGATGAAGAAGACTTCAGGCGCATGAAGACTTTGACTCAG 477  
 QY 1391 CATACAAGACTAATCCAGAGCAAGGACGCTGAAGTGGAGGACTCATTTGATTAAT 1450  
 DB 478 CAGATTAACTTGAGCCCAAGCAGACACCGGTGCTTTGGAATGCTGCTGCAGAGAATT 537  
 QY 1451 CATAAAACGATATGTTCAAGGGAGCTTCGAGACTGGGGTTTGAGCTTTGATTTCAAC 1510  
 DB 538 TCACAAAACGAGACAGCCAGCAATGAGCTGACCGCTGGGGGCTCAGTCTGCAATAAGAT 597  
 QY 1511 TTACTGCTCTTCTCAGGAAGAAATTTTCAAAACAGAAAAGATTCACCAAGGTGGAAAACA 1570  
 DB 598 GTCCACAAGATTGAAGTTCGGCTTCTGCCAATGGAGGATCACTTTAGGAACACTTCA 657  
 QY 1571 TTTGATTACAAATCCAAATTTGCGAGATTGTCACAAAGAAACAAGAGTGCACCAATTAAT 1630  
 DB 658 TTTGT---CACATCGGAGGCGCTGAACTGGGTTAAGAGAGTACAGAGATGCTTCCATT 714  
 QY 1631 AGTGTAAAGCCACTAGATACTGCTGTTGATCTATACCGAAGAAATTTATGAAGAGCC 1690  
 DB 715 CTAATCTATCCATGCTATTTCTGGGCACTCTTTATTCAAAGAGAGCAATGGACCAAGCC 774  
 QY 1691 AATTCTATTGATCAAAAATCTATTAAAGTTTACACAGCCATGGGCATGCAAAATGAGAAA 1750  
 DB 775 AGAAGACTGGTTAACTGTTGGAAGAAAGATTGCGGGCCCATTTGGCATGGCCAAACCC 834  
 QY 1751 GCATATATGATTGAAGT---GGATGACAGAACTGAAGCTTCTTAAGAGTCTTTACAGCAA 1807  
 DB 835 CCAGCTTGGGTTGAGCTGAAGGATGACCGAATAGAGACCTATATCAGGACCAATTCAGTCC 894  
 QY 1808 AAGTCCACAGCAGAC-----ACCCAGATAGTTGCTGCTGTTGTCAAGTAATCGAAG 1861  
 DB 895 TTACTGGGATTGAGGGGAAGATACAAATGGTTCGTTTGCATCATATGGGCACACGTGAT 954  
 QY 1862 GACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAAGTGT 1921  
 DB 955 GATCTCTATGGAGCCATCAAGAGCTGTGCTGCTGCGTCCAGTCCCGCTCAGAGTTC 1014  
 QY 1922 GTGTGGCCCGAACCCTTTAGGCAAAACAGCAAACTGTTCATGGCCATTGCTACAAAGATTGCC 1981  
 DB 1015 ATCAATGTCGAACCATTTGGTTCAGCCACAGGCTTCGGAGCGTGGCTCAGAAAAATTTTA 1074  
 QY 1982 CTACAGATGACATGTCAGAGATGGGAGGAGGAGCTCTGAGGGTGGACATCCCTCGAGCTC 2041  
 DB 1075 CTTCAAGATGAATCTGAACCTGGGTGGTGGAGCTCTGGGAGTGGATATTCGGCTGAACAA 1134  
 QY 2042 GTGATGATCTTGGCATTCGATTTTACCATGATGATGACAGCTGGGGCGGAGGTCAATTCGCA 2101  
 DB 1135 CTAATGCTGATTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1194  
 QY 2102 GGATTTGTTGCCAGCATCAATGAAGGAGTGAACCGCTGTTTCTCAGCTGCGATTAATTTTCA 2161  
 DB 1195 GGCTTCGTGGCCAGCATAAATCTCAGACTCAGCAAAATGGTGTCTCGAGGGTGGTGTTCAG 1254  
 QY 2162 GATAGAGGACAGGAGCTGGTGTAGTGGCTCAAGTCTGCTGCAAGCGGCTCTGAGGGCT 2221

```

Db 1255 ATGCCACATCAGAGATGTTGAGACAGCCGTAAGCTCTGCTGGTGGTTCCTTGAAGAAG 1314
Qy 2222 TGGAAATAGCTGCAATGATGATACATCCCGAGCGGATCATCGTGTACCGGATGCGGTAGA 2281
Db 1315 TATTATGAGGTGACCATGCTCTCCAGAGAAATTTGGTGTATCCGAGATGAGAGTGTCT 1374
Qy 2282 GACGGCGAGCTGAAACATCTGTTGAATACGAGAGGCAAGTTTGGATTTGTATAA 2341
Db 1375 GATGGCGAGCTGAAAGCATGTTCCATCTACGAGATCCCTGAGCTCAGAAAGTGTGAA 1434
Qy 2342 TCCATTGTTAGAGGTTTACCAACCTAGATTAACGTTATTTGGTGAAGAAAGAGTGAAC 2401
Db 1435 GCGTTTGATA---ACTACACACCCCAAGATGATGTTGTGATGTTACAGAAAGAAATCAGC 1491
Qy 2402 ACCAAGTTTTTGGTCTCAGTCTGAGAGAGACTTCAAGATCCACTTCTCGAAGCACTATT 2461
Db 1492 ACCAATCTGTACCTGCTGCTCTCTATCACTTCTGTAACCCCTCCCGGAGCTGTGTT 1551
Qy 2462 GATGTAGAGGTTTACAGACACAGATGATGATGATGATGATGATGATGATGATGATGATG 2521
Db 1552 GATCATACATTAACAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1611
Qy 2522 AGTGTAGTGTGTTCTTCCACACATTAACATGATGATGATGATGATGATGATGATGATG 2581
Db 1612 CAGGGCTGTGAGATACCTACACATGATGATGATGATGATGATGATGATGATGATGATG 1671
Qy 2582 GACCATATACAGCGCTTGAACCTCAACAGCTGTGATGATGATGATGATGATGATGATGATG 2641
Db 1672 GATCATATCAGAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1731
Qy 2642 ATTCGTGTTCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2701
Db 1732 ATCCGAGTCTCAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1791
Qy 2702 ATTCAGAGAGCGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2758
Db 1792 TTGCATCATGAGCGACGACCATCAAGCTGTGTGGAAACCTGTTCTCTGTAACCTGGGA 1848

```

RESULT 8  
AAH15960  
ID AAH15960 standard; cDNA; 2272 BP.

AAH15960;

26-JUN-2001 (first entry)

Human cDNA sequence SEQ ID NO:14558.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-0116126.

29-JUL-1999; 99JP-0248036.

27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI, 2001-318749/34.  
DR  
XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -

Claim 8; SEQ ID 14558; 2537bp + CD ROM; English.

XX  
CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC polynucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

XX  
SQ Sequence 2272 BP; 609 A; 509 C; 585 G; 569 T; 0 other;

Query Match 9.5%; Score 329; DB 22; Length 2272;

Best Local Similarity 51.8%; Pred. No. 6,1e-72;

Matches 871; Conservative 0; Mismatches 791; Indels 18; Gaps 5;

```

Qy 1086 TTGTTCTTACCAATATATACATTAAGCATACAGAGTGTATTTGACTGGACACAGA 1145
Db 34 TTGTTATCACTGATATTAACATGATGATGATGATGATGATGATGATGATGATGATGATG 93
Qy 1146 ATCCCAAGAGACCTTTAAAGAAAGCCGAGCGCTGTGGGCTGACCTCTTAGAATACATCA 1205
Db 94 CTCCAAGAGATGCTTACCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 153
Qy 1206 GGAAGCATCAACCAAGAGATCAACGACTTGAAGAGCGCTGCTTGTGACGACCAAGCCCA 1265
Db 154 GCAGAAATTTATGGATACAGTTAAGAAAGAGACCAAGCATTTGCTGATTCACAGGCCCA 213
Qy 1266 ---AGAGAGCGCGGCGCTGTGGGCGGACACTGCCAGGCGCTGCATGCTCATCTTGAGC 1322
Db 214 GTGAGAGACAGATTAATCATGAGATGCTGCTAAAGAGGGAATCTGCTGCTGCTGAGC 273
Qy 1323 TCTGCTATCTTACAGCTTAAATGATTAATGCTGATATTTTAACGTGATGAAGACT 1382
Db 274 TTCTTTTATGACCGGATCCAGAGAGATGAAGAGGCTTCAAGCCATGAAGATTT 333
Qy 1383 TAGCGTTTATACAGAGCTAATCTCAGAGCAAGGAGCGCTGAGTGAAGTGAAGTGAAGTGA 1442
Db 334 TGGCTCAGCAATTAATCTGAGGCCCAAGCAACATATGCTTTGGAATGCTTGCTGC 393
Qy 1443 ATTACATTCATTAAGAGATGATTAAGAGAGGCTTGAAGCTGAGACTGGGCTTTGAGCTTG 1502
Db 394 AAGAGATTTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 453
Qy 1503 ATTCACTTACTGCTCTCTCAGGAAGATTTTGCAGAAAGAGATTCACCAAGAGT 1562
Db 454 AAGAGATGATCAATTAAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 510
Qy 1563 GAAAGATTTGATTAACATCAATTTTGCAGATTTGCTGCAAGAGAGAGAGAGAGAGAGAGAG 1622
Db 511 AAGATTACTGTTTATCAATCTCAGAGACTAATCTGTTGAGAGAGAGAGAGAGAGAGAGAG 570
Qy 1623 CATTAATTAAGTTAAGCACTAGATTAAGTGTGCTGTTGATCTATACGCGAAGAAATTTAG 1682

```

Db 571 CTTCATCTTGACATATCCCATGCTTTCTGGGCATCTTTTACCCAAAGAGAGCAATGG 630  
QY 1683 AAGCAGCAATTCATTGATACAAAATCTATTTAAAGTTACACAGCCATGGCGATGCAAA 1742  
Db 631 ACCAGGCTCGAGAACTGGTCAACATGTTGGAGAAGATAGCCGCCCCCATTTGGCATGCGTA 690  
QY 1743 TGAGAAAAGCAATAATGATTTGAAGT---GGATGACAGAACTGAAGCCTACTTTAAGAGTCT 1799  
Db 691 TGAGCCCCACCGCGCTGGTGGTGAATTAAGGATGACCGAATAGAGACTTATGTGCAAGAAC 750  
QY 1800 TACAGCAAAAGGTACAGAGAG-----ACACCCAGATAGTGTCTGCTGTTGTCAAGTA 1853  
Db 751 TCAATCCAGCTTAGAGAGCTGAGGGGAAGATACAGATGTTGTTTGCAATCATATGCGGCC 810  
QY 1854 ATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAA 1913  
Db 811 CACGTGATGATCTATATGGGGCCATCAAGAGCTGTGCTGTGTGCAAGTCCCGAGTGCCT 870  
QY 1914 GTCAGTGTGTGGTGGCCGAACTTATAGGCAACAGCAAACTGTTCATGCGCCATGCTACAA 1973  
Db 871 CCCAGGTTGTCAATGTTGCAACCAATTTGGTCAGCCCAACAGGCTTCGAGTGTGGCCCA 930  
QY 1974 AGATTGCCCTACAGATGAATCAAGTCAAGTGGAGAGAGCTCTGGAGGTGGACATCCGCC 2033  
Db 931 AGATTTTACTTCAGATTAATCTGAAATTTGGTGTGAGCTCTGGGAGTGGATATTCCTC 990  
QY 2034 TGAAGCTCGTATGATCGTTGGCATCGATTGTTTACATGACATGACAGCTGGCGGAGGT 2093  
Db 991 TGAACAGTTAATGGTGTATCGGGATGATGTTTACCATGACCCAGTAGAGGATGCGCT 1050  
QY 2094 CAATCGAGGATTTGTTGCGAGATCAATGAAGGATGACCCGCTGGTCTCACCGTGA 2153  
Db 1051 CCGTGTGTGGCTTCTGGGAAGCATCAATCTCACCTTCACAAATGGTATTCCCGGGTGG 1110  
QY 2154 TATTTTCAGGATAGAGCAGGAGCTGTAGATGGGCTCAAGTCTGCCCTGCAAGCGGCTC 2213  
Db 1111 TGTTTCAGATGCGCATGAGGATGTTGGACAGCTGAAAGCTATGCTCTGCTGGGCTCT 1170  
QY 2214 TGAGGCTTTGGAATAGCTGCAATGATGATACATGCCAGCGGATCATCGTGTACCCGATG 2273  
Db 1171 TAAAAAGTTTATGAGGTGAACCACTGTCTACAGAGAAGATTTGGTGTACCGTGTATG 1230  
QY 2274 GGTAGGAGCGGCTGAGTGAACCACTGTGTAACATGAGTGGCAAGTGTGTTTGGAT 2333  
Db 1231 GAGTGTCTGATGGCCCACTGAAGACAGTGTGCCAATATGAGATTCCTCAACTACAGAA 1290  
QY 2334 GTCTAAATCCATTGTTAGAGGTTTACAAACCTAGANTACCGTAAATTTGGTGAAGAAA 2393  
Db 1291 GTTTGAGCTTTTGAGA---ATTATCAGCCCAAGATGGTGTGTTGTTAGTTCAGAAGA 1347  
QY 2394 GAGTGAACACCAAGATTTTTTGTCTGCTGTGGAGGAGACTTCAGAAATCCACTTCTCGAA 2453  
Db 1348 AAATCAGTACTAATCTATATCTGGTGTCTCTCAGAACTTTGTAACCTCCACTCTGGAA 1407  
QY 2454 CAGTATTGATGATGAGTTTACAGACCAAGATGTTATGATGATGATGATGATGATGATG 2513  
Db 1408 CTGTGTGATGATCATACAATAAAGCTGTGAGTGGGTGATTTCTATCTCTTGTGCCATC 1467  
QY 2514 CTGTGAGAAGTGTGATGTTCTTCCACACATTTACAAATGATGATGATGATGATGATGAT 2573  
Db 1468 ATGTACCGCAGGCTGTGCTATCTTACGCAATGATGATGATGATGATGATGATGATGAT 1527  
QY 2574 TGAAGCCAGACACATACAGCGCTTGACCTACAAGCTGTGCCACATCTATTATACACTGCG 2633  
Db 1528 TGAGCCCTCATATGACAGAGGCTGACTTTCAAACTGTGCCACATGATGATGATGATGAT 1587  
QY 2634 CAGGTGTCAATCGTGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2693  
Db 1588 CTGGCACCACATCAGAGTTTCCAGTCTCTTGAAGTATGCTGCTGCTGCTGCTGCTGCTG 1647  
QY 2694 GCCAGATATTACAGAGAGCAAAATCTGTCTACTGTCAACCGGCTTTTACTACCTCTAAC 2753

Db 1648 GACACATCTTGCATCATGACACCCATCCAGCTGTGCGAGAACCTGTTCTTCTCTGTGAC 1707

# RESULT 9

ABL10571  
ID ABL10571 standard; cDNA; 2838 BP.

XX ABL10571;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 26195.  
DE Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX P-PSDB; ABB66468.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

XX Claim 1; SEQ ID NO 26195; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 2838 BP; 783 A; 659 C; 682 G; 714 T; 0 other;

Query Match 8.9%; Score 310.2; DB 23; Length 2838;  
Best Local Similarity 49.1%; Pred. No. 3.5e-67;  
Matches 1133; Conservative 0; Mismatches 1130; Indels 46; Gaps 10;

QY 476 AAAACAGGTTCTTTCAGGCATATAGTAAAGTTAAGCACTAACCAATTCGGCTGACATCC 535

Db 431 AAGAAAGGAGTGGTTGGCACTCATATACCGTGCAGGCAAACTATTTAAGGTATTAAAG 490

QY 536 CGTCCCAGTGGCGCTTATATCATGATATCATATGACTATACCCCACTGATGGAAGCCAGA 595

Db 491 CGTCCAAACTGGACCATCTACCAAGTACCGCGTCAATTTACGCTCATGCTGGAGGTACA 550

QY 596 AGACTCGTTCAGCTCTTCTTTTCAACAGAGATCTAATTTGGAAGTGTATGCTTTT 555

Db 551 CGACTCGACGGTCTTCTTGTATGAACATAAAGGATCCTGGGC--GGCTACATCTTTG 608

QY 656 GATGGAACGATATATTTTACCTAAAGACTACAGCAAA-----AGGTTA 701

Db 609 ACGGAACCAATATGTTTTCATCAATCACTTCAAGCTTCAAGTAGCCCTATGTTT 668





AC AAZ13404;  
 XX 12-OCT-1999 (first entry)  
 XX Human gene expression product cDNA sequence SEQ ID NO:873.  
 DE Human; gene; gene expression product; diagnosis; therapy; probe;  
 KW detection; mapping; tissue typing; profiling; forensic; cancer;  
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.  
 XX Homo sapiens.  
 XX WO9938972-A2.  
 XX 05-AUG-1999.  
 XX 28-JAN-1999; 99WO-US01619.  
 XX 03-APR-1998; 98US-0080666.  
 PR 28-JAN-1998; 98US-0072910.  
 PR 24-FEB-1998; 98US-0075954.  
 PR 31-MAR-1998; 98US-0080114.  
 PR 03-APR-1998; 98US-0080515.  
 XX (CHIR ) CHIRON CORP.  
 PA (HYSE-) HYSEQ INC.  
 XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;  
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;  
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;  
 PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;  
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;  
 XX WPI; 1999-494092/41.  
 DR Novel human genes and their expression products which are  
 PT differentially expressed in different cell types  
 XX Claim 1; Page 862-863; 2479pp; English.  
 XX The present invention describes a library of human polynucleotides  
 CC comprising the sequences given in AAZ12532 to AAZ1779. Also described is  
 CC a method of detecting differentially expressed genes correlated with the  
 CC cancerous state of a mammalian cell, comprising detecting at least one  
 CC differentially expressed gene product in a test sample from a cell  
 CC suspected of being cancerous, where the gene product is encoded by one  
 CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ1779. The  
 CC polynucleotides can be used as a source of primers and probes, which can  
 CC be used for a variety of purpose, e.g. detection of expression levels,  
 CC mapping, tissue typing or profiling, forensics, genetic analysis and  
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides  
 CC can be used for raising antibodies for experimental, diagnostic and  
 CC therapeutic purposes. The polynucleotides may also be used to construct  
 CC arrays for diagnostics (which may be used to determine function of an  
 CC encoded protein); and to detect differences in expression levels between  
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
 CC identify a genetic predisposition or susceptibility to a disease such as  
 CC cancer). The polynucleotides of the invention are especially used in the  
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
 CC and lung cancer. The polynucleotides can also be used to screen for  
 CC peptide analogues and antagonists.  
 XX Sequence 300 BP; 75 A; 70 C; 91 G; 64 T; 0 other;  
 SQ Query Match 8.2%; Score 286.4; DB 20; Length 300;  
 Best Local Similarity 99.3%; Pred. No. 1.1e-61;  
 Matches 298; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
 QY 1909 CCCAAGTCAAGTGTGTGGTCCCGAACCTTAGGCAACAGCAAACTGTCATGGCCATTGC 1968  
 DB 1 CCCAAGTCAAGTGTGTGGTCCCGAACCTTAGGCAACAGCAAACTGTCATGGCCATTGC 60  
 QY 1969 TACAAGATTGCCCTACAGATGAATGCAAGATGGAGGAGAGCTCTGGAGGTGGACAT 2028

Db 61 TACAAGATTGCCCTACAGATGAATGCAAGATGGAGGAGAGCTCTGGAGGTGGACAT 120  
 QY 2029 CCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGCG 2088  
 Db 121 CCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGCG 180  
 QY 2089 GAGGTCAATCGCAGGATTGTTGCCAGCATCAATGAAGGATGACCCGCTGTTCTCAG 2148  
 Db 181 GAGGTCAATCGCAGGATTGTTGCCAGCATCAATGAAGGATGACCCGCTGTTCTCAG 240  
 QY 2149 CTGCATATTTCCAGATAGAGGACAGGAGCTGGTAGATGGCTCAAAGTCTGCCTGCAAGC 2208  
 Db 241 CTGCATATTTCCAGATAGAGGACAGGAGCTGGTAGATGGCTCAGAG-CTGCCTGCAAGC 299  
 RESULT 11  
 ABL10535  
 ID ABL10535 standard; cDNA; 3524 BP.  
 XX ABL10535;  
 AC ABL10535;  
 XX 26-MAR-2002 (first entry)  
 DT 26-MAR-2002 (first entry)  
 XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 26087.  
 DE Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX Drosophila melanogaster.  
 XX WO200171042-A2.  
 XX 27-SEP-2001.  
 XX 23-MAR-2001; 2001WO-US09231.  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX (PEKE ) PE CORP NY.  
 PA Venter JC, Adams M, Li PWD, Myers EW;  
 PI WPI; 2001-656860/75.  
 DR P-PSDB; ABB66432.  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PS Claim 1; SEQ ID NO 26087; 21pp + Sequence Listing; English.  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB57737-AB572072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 3524 BP; 1053 A; 741 C; 761 G; 969 T; 0 other;  
 SQ Query Match 7.5%; Score 259.8; DB 23; Length 3524;  
 Best Local Similarity 47.3%; Pred. No. 1.7e-54;  
 Matches 1085; Conservative 0; Mismatches 1179; Indels 29; Gaps 9;  
 QY 470 GAATCAAAAACAGGTCTTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTG 529



Qy	1606	AGAAACAAGAGGTGACCACTTAAATTAAGTTTAAAGCACTAGATTAACGTGCTGTGATCTTA	1665
Db	1470	AGAAACCAAAAGGATGCTTACCACTCCGAGCGATGGC--CTGATGCTGTGGGCTCATTCGC	1528
Qy	1666	TACGCGAAGAAATTAATGAAGACGCCAAATTCATGTATACAAATCTATTAAAGTTACACC	1725
Db	1529	GCGCGAAAGGAATTCCTCATGAACTCCGAACCTCTACTTGACTCTTTGTATGAGACACTAG	1588
Qy	1726	AGCCATGGGCAATGCAAAATGAGAAAAAC--AAATATGATTGAAGTGAATGACGAATCTGA	1782
Db	1589	TGGAATGGGCTTAAATTTGGAAGCCCCCAGGAATTCATTAATTTATATGATGACACTGG	1648
Qy	1783	AGCCTACTTAAGAGTCTTACACAAAAAGGTCAACAGACACACCAGATAGTTGTCTCT	1842
Db	1649	AACCTTAGTGAAGCAATGGAATGTTGTGTGGCTCAATCCCAACTATTAATATGCTT	1708
Qy	1843	GTGTCAAGTATTCGGAAGGACAAATAGATGCTATTAAAAAATCCTGTGTACAGATTG	1902
Db	1709	GGTACCCAAATGAATAGCGCGAAAGATCACTCAATCAATCAAAAAAGAGAAATACGTTGACG	1768
Qy	1903	CCCAACCCCAAGTCAGATGTGTGGTGGCCCGAACTTTAGGCAAAACAGCAAACCTGCATGGC	1962
Db	1769	GGGCGTGCACCTCA--AGTTGTGACCTTTAAACGACCAAGAACCGTAGCCTTAAG	1825
Qy	1963	CATTGCTAACAAAGTTGCCCCATCAGATGAATCGAAAGTGGAGAGAGCTCTGGAGGTT	2022
Db	1826	CATTGCCACCAAATATAGCAATCCAACTGAATTTGCAAGTTGGGAATATACCCCTGGATGAT	1885
Qy	2023	GGAACATCCCCCTGAAGCTCGTATGATGCTGGGCAATGATTTTAAACATGATGACAGC	2082
Db	1886	GGAACATCCCTGTCCGGAGCTGATGCAATATGGCTTTGACATTTGCGAAGAGACACAGAGA	1945
Qy	2083	TGGCGGAGAGTCAATCGCAGAAATTTGTTCGACGATCAATGAAGGATGACCCGCTGTT	2142
Db	1946	TCGGAAGAGGGCTTAAGGAGCAATGATGCTCTCATGATCTTAACAGCAAACTCCACGTA	2005
Qy	2143	CTCAAGCTGCATTTTACAGATAGAGACAGAGAGCTGTGTATGTGGGCTCAAAAGCTGCTCT	2202
Db	2006	CTTACAGCAAGTCAACGAGTGCAGCGCCCTTTATATGTGCTCGCTTAACACCTTTGGCCGAT	2065
Qy	2203	GCAAGC---GGCTGTGAGGCTCTTGAATATGCTGCATAGATGATGCCCCAGCGATCAT	2259
Db	2066	GATAGCAAAAGGCTCTGGCCCAATATCAATAGACATAGAAAGCTGCATCTCGAATCGT	2125
Qy	2260	CGTGTACCGGATGCGGTAGAGACGCGCAAGCTGAATAACCTGTGAACTACGAATGCC	2319
Db	2126	ATTTTATGAAACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTGAATTTGAAGTCAA	2185
Qy	2320	ACAAGTTTGGATTTGTCTAAATTCATATGTGTAGAGTTTAAACCTTAGANTAAAGTAAT	2379
Db	2186	GGACATCAATTAAGAGTTGAATACTGAATATCGCCGCGCTCAGCTTAACCCACCGCAATT	2245
Qy	2380	TGTGTGTAAGAAAAAGGTGAACACAGATTTTTTGTCAGTGTGAGAGAATTCACGA	2439
Db	2246	AGCT---TATATGTGTGTAACAGATTCATGAACACGCGCTCTTCTCTCAACGACAAAA	2302
Qy	2440	TCCATTTCTCTGGAAACATTTATATGTATAGAGTTACACAGAAATGTATGACTTTT	2499
Db	2303	TCTCTCGGCTGTACTATATGTATATGACGTTAATCTTCCCGGAGATTAACACTTTTA	2362
Qy	2500	TATCGTGAACGAGCTGTGAAGAGTGTATGTCTTCTCCACACATTAACAATGTCATCTA	2559
Db	2363	TCTGTATCTCGAACAAAGTTGTCAAGGTTACAGTGTGCGGACCAACATCAATATGTTCTTTA	2422
Qy	2560	TGACAAACAGCGGCTGAAAGCAGACACATACAGCGCTTGAACCTTCAAGCTGTGCATAT	2619
Db	2423	TAGCAGCAATGGGTCTCTCAACGAGAAAAATGCAAAAATCTTAGTACAAAGATGTGCACTT	2482
Qy	2620	CAATTAAACAATGGCAGAGTGCATTCGTGTCTCTGCTCTTGTCCAGTACGCCACAACT	2679
Db	2483	GTACTCAATTTGTGTGGGACACACAGAGTGCAGAGCTTTCAGATGAGCTTAAGAACT	2542



QY 1666 TACGGAGAAATTAAGAGAGCAATTCATGATACAAATCTATTAAAGTTACACC 1725  
 DB 1529 GCGGCAAAAGAAATCCAGATGAACTCGAACTCTTGTATGAGAGAGCTAG 1588  
 QY 1726 AGCATGGGATGCAAAATGAGAAAAGC--ATAATGATGAGAGTGAAGACATGA 1782  
 DB 1589 TGGAAATGGGCTTAGAATTCGAAGCCCCAGGAATTCATTAATTAATGATGCGACCTGG 1648  
 QY 1783 AGCTTACTTAAGAGCTTACAGCAAAAGTTCACAGACAGCCGAGATGTCGTCT 1842  
 DB 1649 AACTTATGAGAGCAATGATGATGTCGCTGACATGATCCCAATTAATATGCTT 1708  
 QY 1843 GTTGTAAAGTATCGGAAGAGCAATATGATGCTATTAAAAATTAATCTGTGATGAT 1902  
 DB 1709 CGTACCAATGATAGCGCGAAAGATACATCAATCAAAAAGAGATAGTTGACAG 1768  
 QY 1903 CCGTACCCCAAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1962  
 DB 1769 GCGGATGCTCAACTCA--AGTTGTGACCTTTAAACGACCAAGAACCTGATGAG 1825  
 QY 1963 CATTGCTACAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2022  
 DB 1826 CATTGCTACAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1885  
 QY 2023 GGAATCCCTGAAAGCTGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2082  
 DB 1886 CGAACTACCTTGTCCGAGCTGATGACAAATGCTTGTGATGCTGCTGCTGCTGCTGCT 1945  
 QY 2083 TGGGCGAGAGCTCAATCCGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2142  
 DB 1946 TCGGAAGAGGCTCAAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2005  
 QY 2143 CTCAAGCTGATATTTAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2202  
 DB 2006 CTTCAGGACAGTCAAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2065  
 QY 2203 GCAAGC--GGCTGAGAGGCTTGGAAATGCTGATGATGATGATGATGATGATGATGAT 2259  
 DB 2066 GATACCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2125  
 QY 2260 CGTGTACGCGATGCTGATGAGAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2319  
 DB 2126 ATTTTATGAGAGCGGTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2185  
 QY 2320 AAGATTTTGTGATGCTTAAATCATTGATGATGATGATGATGATGATGATGATGATGAT 2379  
 DB 2186 GGAATCATTTGAGAGAGTTGAAAAGTAAAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 2245  
 QY 2380 TGTGTGAGAAAGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 2439  
 DB 2246 AGCT--TATATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2302  
 QY 2440 TCCATTTCTGGAACGATTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2499  
 DB 2303 TCTTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2352  
 QY 2500 TATCGGAGCGAGGCTGATGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2559  
 DB 2363 TCTGTGCTGCAAGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2422  
 QY 2560 TGAACAAGCGGCTGATGAGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 2619  
 DB 2423 TAGAGAGATGAGTGTCTCAACCGAGAAATGCAAAAATTAAGTAAAGTAAAGTAAAGTAAAG 2482  
 QY 2620 CTATTAAATGCGCGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2679  
 DB 2483 GTACTACAAATGCTGCGGACCAAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2542  
 QY 2680 GAGCTTTCTGTTGCGCAAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAAACCGCT 2739  
 DB 2543 AGCTACCTCGTGGGTACGAATCTGATCTATTCGCAAAAACGCGCTGAAAAGAGT 2602  
 QY 2740 TTACTACTCTAA 2752

DB 2603 TTATTATCTATTA 2615

# RESULT 13

ID AAH14032 standard; cDNA; 1733 BP.

AAH14032;

26-JUN-2001 (first entry)

Human cDNA sequence SEQ ID NO:11143.

Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-0116126.

29-JUL-1999; 99JP-0248036.

27-AUG-1999; 99JP-0300253.

11-JAN-2000; 2000JP-0118776.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

WPI; 2001-318749/34.

Claim 8; SEQ ID 11143; 2537bp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602  
 full-length cDNAs defined in the specification. Where a primer set  
 comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 to the complementary strand of a polynucleotide which comprises one of  
 the 5602 nucleotide sequences defined in the specification, where the  
 oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 of an oligonucleotide comprising a sequence complementary to the  
 complementary strand of a polynucleotide which comprises a 5'-end  
 sequence and an oligonucleotide comprising a sequence complementary to a  
 polynucleotide which comprises at least 15 nucleotides and the combination of  
 the 5'-end sequence/3'-end sequence can be selected from those defined in  
 the specification. The primer sets can be used in antisense therapy and  
 in gene therapy. The primers are useful for synthesizing polynucleotides,  
 particularly full-length cDNAs. The primers are also useful for the  
 detection and/or diagnosis of the abnormality of the proteins encoded by  
 the full-length cDNAs. The primers allow obtaining of the full-length  
 cDNAs easily without any special methods. AAH03166 to AAH13628 and  
 AAH13633 to AAH18742 represent human cDNA sequences; AAH9246 to  
 AAH9589 represent human amino acid sequences; and AAH13629 to AAH13632  
 represent oligonucleotides, all of which are used in the exemplification  
 of the present invention.

Sequence 1733 BP; 435 A; 408 C; 451 G; 439 T; 0 other;

Query Match 6.9%; Score 238.8; DB 22; Length 1733;  
 Best Local Similarity 52.1%; Pred. No. 2.2e-49;  
 Matches 609; Conservative 0; Mismatches 548; Indels 12; Gaps 3;

QY	1594	AGATTGGTCCAAAGAAACAAGAGGTGCACCAATTAATTAAGTGTTAAGCCACTAGATAACTG	1653
Db	3	AAACTGGGTTAAGGAAGTAACCCAGAGACCCCTTCCATCTTGACTATCCCATGCAATTTCTG	62
QY	1654	GCTGTTGATCTATACGCGAAGAAATTATGAAGCAGCCAAATTCATTGATACAAAAATCTATT	1713
Db	63	GGCACTTTTATACCCAAAGAGAGCAATGGACCCAGGCTCGAAGACTGTCACATGTTTGG	122
QY	1714	TAAAGTTACACAGCCATGGCGATCGCAAAATCAGAAAGCAATATGATGATTCGAAGT---	1770
Db	123	GAAGATAGCCGGCCCATTTGCGATGCGTATGAGCCCAACCGCCCTGGGTGTAATAAGGA	182
QY	1771	TGACAGAACTAAGCCTACTTAAAGATGCTTACAGCAAAAGGTCACAGCAG-----ACAC	1824
Db	183	TGACCGAATAGAGACTTATGTGCAAGACCAATTCATATCCAGCTTAGGAGCTGAGGGGAAT	242
QY	1825	CCAGATAGTTGTCGTCTGTTGTCACAGTAATCCGAGAGCAAAATACGATGCTATTAAAA	1884
Db	243	ACAGATGGTTGTTTGCATCATATGGGCCCAAGTGATGATCTCTATGGGGCCCATCAATAA	302
QY	1885	ATACCTGTGTACAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCCAA	1944
Db	303	GCTGTGCTGTGTGCAAGTCCCGAGTGCCTCCAGGTTGTCAAGTTTCGAACCAATTGGTCA	362
QY	1945	ACAGCAAACTGTCAATGSCCATGCTACAAAGATTGCCCTACAGATGCAATCGCAAGTGG	2004
Db	363	GCCACCAGGCTTCGGAGTGTGCCCCAGAAGATTTTACTTCAGATTAACGTGTAATTTGG	422
QY	2005	AGGAGACTCTGAGGGTGGACATCCCTCGAAGTCGTGATGATCGTTGGCATCGATTG	2064
Db	423	TGTTGAGCTCTGGGAGTGGATATTCCTCTGAAACAGTTAATGGTGTATCGGATGGATGT	482
QY	2065	TTACCATGACATCACAGCTGGGGAGGTCAATCCGACAGATTGTTGCCAGCATCAATGA	2124
Db	483	TTACCATGACCCAGTAGAGGCAATGGCCTCCGTGGTGGCTTTCTGGCAAGCATCAATCT	542
QY	2125	AGGGATGACCGGCTGTTCTCACTGTCATATTTTCCAGATAGACACAGGAGCTGGTAGA	2184
Db	543	CACCTTCACAAAATGTTATTCCTGGGTGGTGTTCAGATGCCGATCAGAGATTGTGGA	602
QY	2185	TGGGCTCAAAGTCTGCTGCACAGGGCTCTGAGGGCTTGGATAGCTGCATAGTAGTACAT	2244
Db	603	CAGCCTGAAGCTATGCTCTGGGCTCTTAAAAAGTTTATGAGGTGAAACCACTGTCT	662
QY	2245	GCCAGCCGGATCATCGTGTACCGCGATGGCGTAGGAGACGCCAGCTGAAAAACACTGGT	2304
Db	663	ACCAGAGAAGATTGTGGTGTACCGTGTATGGAGTGTCTGATGCCCACTGAAGACAGTTGC	722
QY	2305	GAACTACGAAGTGCCACAGTTTTCGATTGTCTAAATTCATTTGGTAGAGGTTACAAACC	2364
Db	723	CAACTATGAGATTCTCTCAACTACAGAAGTGTTTTGAAGCTTTTGAGA---ATTATCAGCC	779
QY	2365	TAGANTAAACGTAATTGTGGTGAAGAAAGAGTGAACACCAAGATTTTTTGTCTCAGTCTGG	2424
Db	780	CAAGATGGTGGTGTGTGTAGTTCAGAGAAATCAGTACTTAATCTATCTATCTGGCTGCTCC	839
QY	2425	AGGAAGACTTCAGAAATCCACTTCTCTGGNACAGTTATTGATGTAGAGGTTACCAGACAGA	2484
Db	840	TCAGAACTTTGTAACTCCCACTCTCTGAACTGTGGTAGATCATACAATAACAAGCTGTGA	899
QY	2485	ATGATAGACTTTTTTATCTGTGAGCCAGGCTGTGAGAAGTGGTAGTGTCTTCCCAACA	2544
Db	900	GTGGGTGGATTCTATCTTCTTGCCCATCATGTACGCCAGGGCTGTGGCAATCTTAGCA	959
QY	2545	TTACAATGTCACTATGACAAACAGCGGCTTGAAGCCAGACCAATACAGCGTTGACCTTA	2604
Db	960	TTATGTCTGTGTTCTCAACACCGCAAACTTGAGCCCTGATCATATGACAGGCTGACTTT	1019
QY	2605	CAAGCTGTGCCACATCTATTACAACGTGCCAGGTGTCAATTCGTGTTCTCTGCTCCCTGCA	2664
Db	1020	CAAACTGTGCCACATGTAAGTAAATGGCCCTGGCAACATCAGAGTTCCAGCTCCTTTGCA	1079
QY	2665	GTAGCCCAACAGCTGGCTTTTCTTGTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCT	2724

```

Db 1080 GTATGCCCAACAGCTAGCTTTCTCTGCAGGACACATCTTGCATCATGACACGACCATCA 1139
QY 2725 ACTGTCAACCGCCTTTACTACCTCTAAC 2753
Db 1140 GCTGTGCGGAACCTGTCTCTCTCTGTGAC 1168

RESULT 14
ABL10570
ID ABL10570 standard; cDNA; 5671 BP.
XX AC ABL10570;
XX DT 26-MAR-2002 (first entry)
XX DE
XX KW Drosophila melanogaster expressed polynucleotide SEQ ID NO 26192.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX FR 11-JUL-2000; 2000US-0614150.
XX PA (PEXE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR P-PEDB; ABB66467.
XX PT
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions
XX PS Claim 1; SEQ ID NO 26192; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 5671 BP; 1723 A; 1127 C; 1131 G; 1690 T; 0 other;

Query Match 5.3%; Score 184.6; DB 23; Length 5671;
Best Local Similarity 52.8%; Pred. No. 1.6e-35;
Matches 422; Conservative 0; Mismatches 375; Indels 3; Gaps

QY 1956 TCATGCCCAATTCCTACAAGATTGCCCTACAGATGAATCGCAAGATGGAGAGCTCT 2015
Db 3759 TGATGTGATGCCCAAGGTGTATTCAATGAACCCAAATTGATGGAGCTCCCT 3818
QY 2016 GGAGGTGACATCCCCCTGAAGCTGTGATGATCTGGCATGCTATTGTTCACATGACA 2075
Db 3819 GGCAGGTAGTATCCCCCTCCACGGTCTGATGACTGTGGTTTGATGTCTGCCATTCAC 3878
QY 2076 TGACAGCTGGCGGAGTCAATCGCAGGATTTGTTGCCAGCAT---CAATGAAGGGATGA 2132

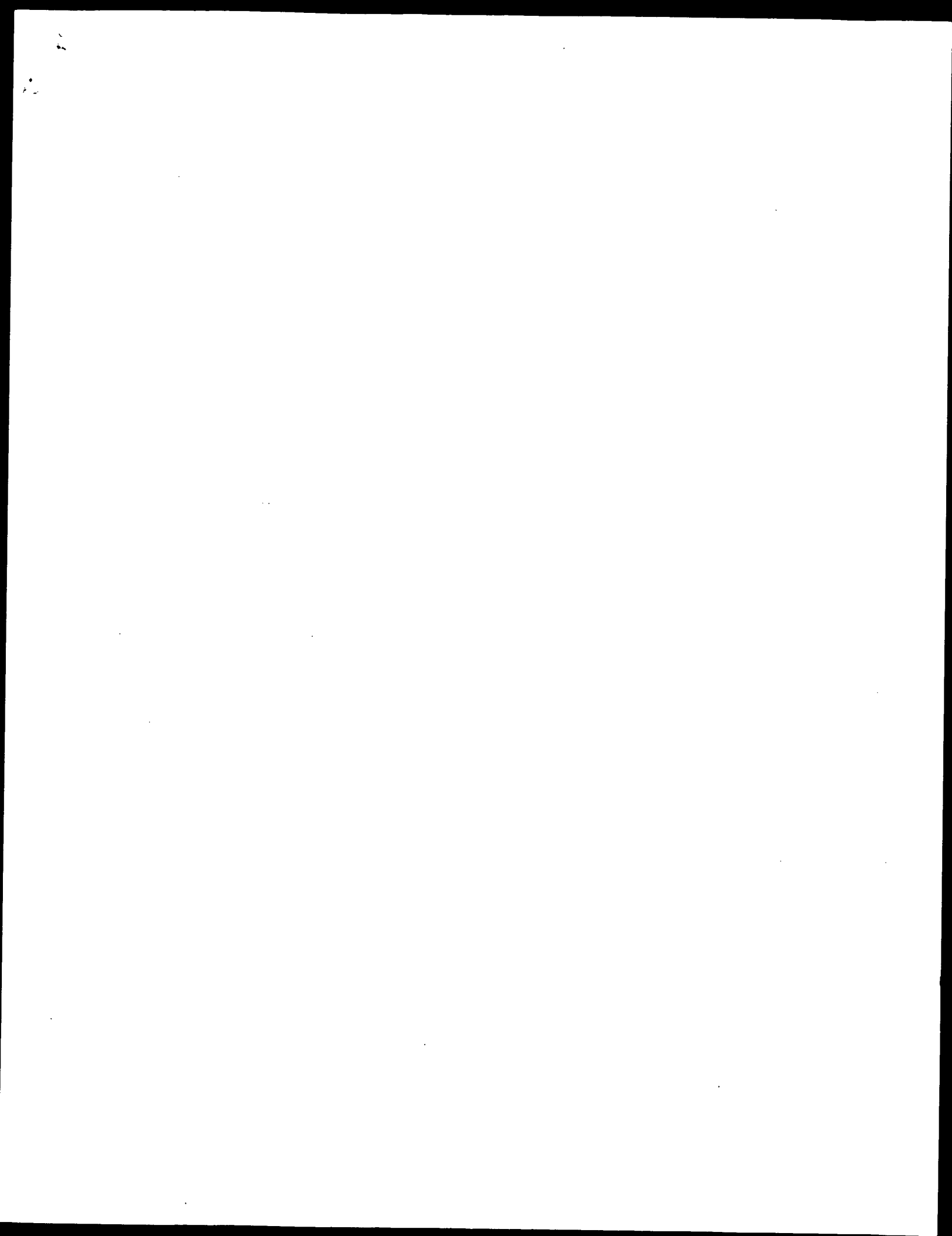
```

Db 3879 CGAAGATTAAGAACAGTCATATGCGGCTTTGTTGCAACCATGAGACCAAGAGACTTT 3938  
QY 2133 CCGGCTGGTCTCAGCGTCGATATTTTTCAGATAGAGACAGAGCTGTAGATGGGCTCA 2192  
Db 3939 TCCGGTACTTCTCCACCGTAAACAGACATTTAGGGCCAGAGATTTTCCGAGAGATGT 3998  
QY 2193 AAGTCTGCTGACAGCGGCTCTGAGGGCTTGGATAGCTGCATAGATCAATGCCACCC 2252  
Db 3999 CCGTAAACATGCGCTGCGCCCTGAGGTGTATCAGAGCAACACCGTCTTTTGGCAGAGC 4058  
QY 2253 GCATATCTGTATACCGGATGCGGTAGAGAGACGCGCAGCTGAAAACTGTGTGAATACG 2312  
Db 4059 GCATTTCTTCTTTCGCGAGCGGTGTGGCGATGTGTCTTACAGGTGGTAAACAGCG 4118  
QY 2313 AAGTGCACAGTTTGTGATTTGTCTAAATCCATTGTGTAGAGTTTCAACCTTGANTAA 2372  
Db 4119 AGGTGAACACCTTAAAGACAGGCTCGACGAAATTTACAAATCAGCTGGCAACAGAGG 4178  
QY 2373 CCGTAAATTTGTGTAAGAAAAGAGTGAACACAGATTTTGTCTGAGTGTGAGAAAGAC 2432  
Db 4179 GCTGTGCAATGACATTTATTAATTTGATTCAGCGCATTAATTTCTGCTACTTTACTGGC 4238  
QY 2433 TTGAGATCCACTTCTCTGGAACAGTATTTGATGATGATGATGATGATGATGATGATG 2492  
Db 4239 ATGCAACCCAGTTCCGGGCACTGTATGATGATGATGATGATGATGATGATGATGATG 4298  
QY 2493 ACTTTTATTCGTGAGCCAGGCTGTGAGAAAGTGTGTGTGTCTTCCACATTTCAATG 2552  
Db 4299 ACTTCTTCTAGTGTCCAGGCTGTGTGAGTGAAGTGTGTGTGTGTGTGTGTGTGTGTG 4358  
QY 2553 TCATCTATGACACAGCGGCTGGAAGCAGACACATACAGGCTGTGAGCTCAAGCTGT 2612  
Db 4359 TTAATTTCTGCAACATGAGGATTAACCGGATTAAGCTGAGATGTCTCTTAAGATGA 4418  
QY 2613 GCCACATCTATTAACAAGTGCAGGTGTCAATGTTCTGCTCTTCCGATAGCCCC 2672  
Db 4419 CCGATATGATCTCAATTAACAGGGAACATACAGATCCCGCTGTCTGCTCAATAGGCC 4478  
QY 2673 ACAAGCTGCTTTTCTTGTGTGGCCAGAGTATTCACAGAGAGCAATCTGTCACTGCA 2732  
Db 4479 ACAATTTGGCTTCTCTGTGAGCGAATCCATTAATGCTGCGCTTTCAGCAGAGCTGCA 4538  
QY 2733 ACCGCTTTACTACTCTTAA 2752  
Db 4539 ATCAATTTACTTTTGTAA 4558

RESULT 15  
ABN62610  
ID ABN62610 standard; cDNA; 548 BP.  
AC ABN62610;  
XX 28-JUN-2002 (first entry)  
DE Human cancer related polynucleotide SEQ ID NO 2577.  
XX Human; cytostatic; gene expression; gene mapping; tissue profiling;  
KM gene therapy; cancer; tumour; gene; ss.  
XX Homo sapiens.  
OS WO200214500-A2.  
XX 21-FEB-2002.  
PD 16-AUG-2001; 2001WO-US25840.  
PF 16-AUG-2000; 2000US-226326P.  
PR 16-AUG-2000; 2000US-226326P.  
XX (CHIR) CHIRON CORP.  
PA (HISE-) HISEQ INC.  
XX

PI Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;  
PI Lanson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;  
XX WPI; 2002-241905/29.  
DR  
XX New nucleic acid for producing a polypeptide, detecting differentially  
XX expressed genes correlated with a cancerous state of a mammalian cell,  
PT and inhibiting tumor growth -  
XX  
XX Claim 1; SEQ ID NO 2577; 883bp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated polynucleotide (ABN27253-ABN33262)  
CC with cytostatic activity. The polynucleotide is used to produce a  
CC polypeptide, to detect differentially expressed genes correlated with a  
CC cancerous state of a mammalian cell and to inhibit tumor growth. The  
CC polynucleotide is used as a probe in mapping and tissue profiling. The  
CC encoded polypeptide and antibodies to the polypeptide can also be used  
CC for therapeutic and diagnostic purposes. The polynucleotide is useful for  
CC gene therapy.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 548 BP; 151 A; 116 C; 146 G; 134 T; 1 other;  
SQ  
Query Match 4.9%; Score 170.8; DB 24; Length 548;  
Best Local Similarity 57.0%; Pred. No. 1,4e-32;  
Matches 310; Conservative 0; Mismatches 234; Indels 0; Gaps 0;  
QY 1943 AACACGAACTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2002  
Db 5 AACACGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 64  
QY 2003 GAGAGAGCTCTGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2062  
Db 65 GAGAGGAGCTGTGGGCTGTGAAATTAAGTCCGATGTGTGTGTGTGTGTGTGTGTGTGTG 124  
QY 2063 TGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2122  
Db 125 GTCTGTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 184  
QY 2123 GAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2182  
Db 185 CCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 244  
QY 2183 GATGGCTCAAGTGTGCTGTGCAAGCGCTCTGAGGCTTGGAAATAGCTGCAATGATGATG 2242  
Db 245 GATGCTTGAAGTTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 304  
QY 2243 ATGCCAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2302  
Db 305 TTGCCAGCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 364  
QY 2303 GTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2362  
Db 365 ATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 424  
QY 2363 CTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2422  
Db 425 TCAAGCTGT 484  
QY 2423 GAGAGAGAGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2482  
Db 485 AACGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 544  
QY 2483 GAAT 2486  
Db 545 GAAT 548

Search completed: January 9, 2003, 19:04:28  
Job time : 504 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 18:55:10 ; Search time 88 Seconds  
(without alignments)  
12099.792 Million cell updates/sec

Title: US-09-873-737A-5

Perfect score: 3472  
Sequence: 1 gggcggaattggggccgacgt.....taaataaatgatgtcc 3472

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:  
1: /cgn2\_6/pdata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/pdata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/pdata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/pdata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/pdata/1/ina/PCUS.COMB.seq:\*  
6: /cgn2\_6/pdata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60.8	1.8	2365	4 US-09-249-697A-5	Sequence 5, Appl
2	60.8	1.8	2365	4 US-09-249-697A-18	Sequence 18, Appl
3	60.8	1.8	2365	4 US-09-363-316B-5	Sequence 5, Appl
4	60.8	1.8	2365	4 US-09-363-316B-23	Sequence 23, Appl
5	57.6	1.7	7218	1 US-08-232-463-14	Sequence 14, Appl
6	57	1.6	1613	3 US-08-838-830-28	Sequence 28, Appl
7	52.8	1.5	2217	4 US-09-244-314-1	Sequence 1, Appl
8	52.8	1.5	2217	4 US-09-498-959-1	Sequence 1, Appl
9	52.4	1.5	7218	1 US-08-232-463-14	Sequence 14, Appl
10	51.8	1.5	19124	2 US-08-487-826B-13	Sequence 13, Appl
11	51	1.5	797	1 US-08-592-126-73	Sequence 73, Appl
12	50.4	1.5	3388	4 US-09-141-206-1	Sequence 1, Appl
13	50.2	1.4	4203	2 US-08-866-757-1	Sequence 1, Appl
14	50.2	1.4	4203	2 US-08-866-757-1	Sequence 1, Appl
15	50	1.4	658	4 US-09-153-593-1	Sequence 1, Appl
16	50	1.4	5852	1 US-07-867-106-2	Sequence 595, App
17	49.4	1.4	1230	4 US-09-387-574-5	Sequence 5, Appl
18	49.4	1.4	1230	4 US-09-668-096-5	Sequence 5, Appl
19	49.2	1.4	2447	2 US-08-628-417-6	Sequence 6, Appl
20	48.8	1.4	2447	2 US-09-014-969-14	Sequence 14, Appl
21	47	1.4	3396	4 US-08-989-299-1	Sequence 1, Appl
22	45.6	1.3	2251	4 US-08-991-677-11	Sequence 11, Appl
23	45.6	1.3	6243	2 US-09-056-075-1	Sequence 1, Appl
24	45.6	1.3	6124	4 US-08-213-419B-3	Sequence 3, Appl
25	44.8	1.3	163	4 US-09-182-145-51	Sequence 51, Appl
26	44.8	1.3	163	4 US-09-182-145-51	Sequence 51, Appl
27	44.8	1.3	488	4 US-08-991-789A-180	Sequence 180, App

28	44.8	1.3	488	4 US-09-062-451-180	Sequence 180, App
29	44.8	1.3	488	4 US-09-598-326-180	Sequence 180, App
30	44.8	1.3	507	4 US-08-991-789A-182	Sequence 182, App
31	44.8	1.3	507	4 US-09-062-451-182	Sequence 182, App
32	44.8	1.3	507	4 US-09-598-326-182	Sequence 182, App
33	44.8	1.3	1222	4 US-08-861-774E-83	Sequence 83, Appl
34	44.8	1.3	2064	4 US-09-178-109-3	Sequence 3, Appl
35	44.8	1.3	2121	4 US-09-004-838-93	Sequence 93, Appl
36	44.8	1.3	5506	4 US-09-004-838-93	Sequence 93, Appl
37	44	1.3	44	1 US-08-381-572-5	Sequence 5, Appl
38	44	1.3	44	1 US-08-381-572-5	Sequence 5, Appl
39	44	1.3	44	1 US-08-592-820-6	Sequence 6, Appl
40	44	1.3	44	1 US-08-592-820-6	Sequence 6, Appl
41	44	1.3	44	2 US-08-670-707A-13	Sequence 13, Appl
42	44	1.3	44	4 US-09-037-601-13	Sequence 5, Appl
43	44	1.3	44	4 US-09-234-393-5	Sequence 19, Appl
44	44	1.3	44	4 US-09-124-238A-19	Sequence 23, Appl
45	44	1.3	44	4 US-09-323-873A-23	Sequence 23, Appl

## ALIGNMENTS

```

RESULT 1
US-09-249-697A-5
Sequence 5, Application US/09249697A
Patent No. 6392018
GENERAL INFORMATION:
APPLICANT: Ford, John
APPLICANT: Yeung, George
TITLE OF INVENTION: NOVEL EGF MOTIF PROTEIN OBTAINED FROM A CDNA LIBRARY OF FETAL LIVER SPLEEN
FILE REFERENCE: 24011-727
CURRENT APPLICATION NUMBER: US/09/249,697A
PRIOR FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 08/968,800
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 5
LENGTH: 2365
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (205)...(1866)
US-09-249-697A-5

Query Match      1.8%; Score 60.8; DB 4; Length 2365;
Best Local Similarity 90.3%; Pred. No. 9.7e-07;
Matches 65; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 58 GATTCATCTAATACGACGACTATAGGCTCGAGCGGCCGCCGCGACAGTGTGAGGT 117
DB 7 GATTCATCTTAATACGACGACTATAGGCTCGAGCGGCCGCCGCGACAGTGTGAGG 66

QY 118 GCAAGACCCAGG 129
DB 67 GACAGACACCCGG 78

RESULT 2
US-09-249-697A-18
Sequence 18, Application US/09249697A
Patent No. 6392018
GENERAL INFORMATION:
APPLICANT: Ford, John
APPLICANT: Yeung, George
TITLE OF INVENTION: NOVEL EGF MOTIF PROTEIN OBTAINED FROM A CDNA LIBRARY OF FETAL LIVER SPLEEN
FILE REFERENCE: 24011-727
CURRENT APPLICATION NUMBER: US/09/249,697A
PRIOR FILING DATE: 1999-02-12

```



; PRIOR APPLICATION NUMBER: US 08/968,800  
; PRIOR FILING DATE: 1997-11-22  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 18  
; LENGTH: 2365  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (205)..(1863)  
US-09-249-697A-18

Query Match 1.8%; Score 60.8; DB 4; Length 2365;  
Best Local Similarity 90.3%; Pred. No. 9.7e-07;  
Matches 65; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 58 GATTCATCTCTAATACGACTCACTATAGGCTCGAGCGCGCCGCGGCGAGGTCTGAGGT 117  
DB 7 GATTCATCTCTAATACGACTCACTATAGGCTCGAGCGCGCCGCGGCGAGGTCTGAGGT 66

QY 118 GCAAGGACCAGG 129  
DB 67 GACAGCACCCGG 78

## RESULT 3

US-09-363-316B-5  
; Sequence 5, Application US/09363316B  
; Patent No. 6392019  
; GENERAL INFORMATION:  
; APPLICANT: Ford, John  
; TITLE OF INVENTION: EGF MOTIF PROTEIN MATERIALS AND METHODS  
; FILE REFERENCE: 28110/35852  
; CURRENT APPLICATION NUMBER: US/09/363,316B  
; PRIOR FILING DATE: 1999-07-28  
; PRIOR APPLICATION NUMBER: US 09/249,697  
; PRIOR FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 08/968,800  
; PRIOR FILING DATE: 1997-11-22  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 2365  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (205)..(1866)  
; NAME/KEY: misc feature  
; LOCATION: (357)  
; OTHER INFORMATION: Xaa = any amino acid  
US-09-363-316B-5

Query Match 1.8%; Score 60.8; DB 4; Length 2365;  
Best Local Similarity 90.3%; Pred. No. 9.7e-07;  
Matches 65; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 58 GATTCATCTCTAATACGACTCACTATAGGCTCGAGCGCGCCGCGGCGAGGTCTGAGGT 117  
DB 7 GATTCATCTCTAATACGACTCACTATAGGCTCGAGCGCGCCGCGGCGAGGTCTGAGGT 66

QY 118 GCAAGGACCAGG 129  
DB 67 GACAGCACCCGG 78

## RESULT 4

US-09-363-316B-23  
; Sequence 23, Application US/09363316B  
; Patent No. 6392019  
; GENERAL INFORMATION:

; APPLICANT: Ford, John  
; APPLICANT: Yeung, George  
; TITLE OF INVENTION: EGF MOTIF PROTEIN MATERIALS AND METHODS  
; FILE REFERENCE: 28110/35852  
; CURRENT APPLICATION NUMBER: US/09/363,316B  
; CURRENT FILING DATE: 1999-07-28  
; PRIOR APPLICATION NUMBER: US 09/249,697  
; PRIOR FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 08/968,800  
; PRIOR FILING DATE: 1997-11-22  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 23  
; LENGTH: 2365  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (205)..(1863)  
US-09-363-316B-23

Query Match 1.8%; Score 60.8; DB 4; Length 2365;  
Best Local Similarity 90.3%; Pred. No. 9.7e-07;  
Matches 65; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 58 GATTCATCTCTAATACGACTCACTATAGGCTCGAGCGCGCCGCGGCGAGGTCTGAGGT 117  
DB 7 GATTCATCTCTAATACGACTCACTATAGGCTCGAGCGCGCCGCGGCGAGGTCTGAGGT 66

QY 118 GCAAGGACCAGG 129  
DB 67 GACAGCACCCGG 78

## RESULT 5

US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 7218 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 IMMEDIATE SOURCE:  
 CLONE: PTZgp-F1s  
 US-08-232-463-14

Query Match 1.7%; Score 57.6; DB 1; Length 7218;  
 Best Local Similarity 3.1%; Pred. No. 1,5e-05;  
 Matches 12; Conservative 227; Mismatches 152; Indels 0; Gaps 0;

QY 104 GGAGGCTGAGGTGCAAGACCAAGACGAGGCGGAGGCGGCTCCAGAAATAGAAA 163  
 DB 1434 RRR 1375  
 QY 164 ACAATGACTGGAGAGCCCGAGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 223  
 DB 1374 RRR 1315  
 QY 224 CAGCTGTGGGCTCCAGCTCCAGCACTGTTATATTCAGCTAGGCTCAGCG 283  
 DB 1314 RRR 1255  
 QY 284 CCACGAGCAGAGGAGGAAATATTTGGCGTGAGCGGAGAGGAGGAGGAGGAGG 343  
 DB 1254 RRR 1195  
 QY 344 GCCAGTCAAGAGCTCCAGATATCTGCTGATTCAGAGTTCGNTAGCAGAGG 403  
 DB 1194 RRR 1135  
 QY 404 GAGAGTGTGTGAGATTTTCATGATCTTGATGATATACAGGAGGAGGAGGAGG 463  
 DB 1134 RRR 1075  
 QY 464 GTTAAGATCAAAAGAGGTTCTTCAGGCA 494  
 DB 1074 RRRRRRRATCGCAAGCTCCCTCGAGCTGCA 1044

RESULT 6  
 US-08-938-830-28  
 Sequence 28, Application US/08938830  
 Patent No. 6040437  
 GENERAL INFORMATION:  
 APPLICANT: Lasky, Laurence A.  
 APPLICANT: Dowbenko, Donald J.  
 TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage  
 TITLE OF INVENTION: Furorow-Associated Proteins (PSPRPs)  
 NUMBER OF SEQUENCES: 73  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Genentech, Inc.  
 STREET: 1 DNA Way  
 CITY: South San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94080  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5 inch, 1.44 MB floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Winpatin (Genentech)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/938,830  
 FILING DATE:  
 CLASSIFICATION: 530  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/798419  
 FILING DATE: 07-FEB-1997  
 ATTORNEY/AGENT INFORMATION:

NAME: Dreyer, Ginger R.  
 REGISTRATION NUMBER: 33,055  
 REFERENCE/DOCKET NUMBER: P1066P1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 650/225-3216  
 TELEFAX: 650/952-9881  
 INFORMATION FOR SEQ ID NO: 28:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1613 base pairs  
 TYPE: Nucleic Acid  
 STRANDEDNESS: Single  
 TOPOLOGY: Linear  
 US-08-938-830-28

Query Match 1.6%; Score 57; DB 3; Length 1613;  
 Best Local Similarity 81.5%; Pred. No. 8.7e-06;  
 Matches 66; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 46 CCGCGGAATTCGATTCATCTAATACACTCACTATAGGCTCCGAGCGCCCGGG 105  
 DB 67 CTGCAAAATTCGGCTTCATCTAATACACTCACTATAGGCTCCGAGCGCCCGGG 126  
 QY 106 CAGCTGTAGGTGCAAGACC 126  
 DB 127 CAGCTGTAGATTCAGCGCC 147

RESULT 7  
 US-09-244-314-1  
 Sequence 1, Application US/09244314  
 Patent No. 6274362  
 GENERAL INFORMATION:  
 APPLICANT: Hodge, Martin R.  
 APPLICANT: Yowe, David  
 TITLE OF INVENTION: RGS-Containing Molecules and Uses Thereof  
 FILE REFERENCE: 5800-19, 035800/174680  
 CURRENT APPLICATION NUMBER: US/09/244,314  
 NUMBER OF SEQ ID NOS: 4  
 CURRENT FILING DATE: 1999-02-04  
 SOFTWARE: Patentin Ver. 2.0  
 SEQ ID NO 1  
 LENGTH: 2217  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (160)..(867)  
 US-09-244-314-1

Query Match 1.5%; Score 52.8; DB 4; Length 2217;  
 Best Local Similarity 89.1%; Pred. No. 0.00016;  
 Matches 57; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 51 GAATTCGATTCATCTAATACACTCACTATAGGCTCCGAGCGCCCGGGCAGGT 110  
 DB 1 GAATTCGATTCATCTAATACACTCACTATAGGCTCCGAGCGCCCGGGCAGGT 60  
 QY 111 CTGA 114  
 DB 61 ATAA 64

RESULT 8  
 US-09-498-959-1  
 Sequence 1, Application US/09498959  
 Patent No. 6410240  
 GENERAL INFORMATION:  
 APPLICANT: Hodge, Martin R.  
 APPLICANT: Yowe, David  
 TITLE OF INVENTION: RGS-Containing Molecules and Uses  
 TITLE OF INVENTION: Thereof  
 FILE REFERENCE: 5800-19A  
 CURRENT APPLICATION NUMBER: US/09/498,959

;; CURRENT FILING DATE: 2000-02-04  
;; EARLIER APPLICATION NUMBER: 09/244,314  
;; EARLIER FILING DATE: 1999-02-04  
;; NUMBER OF SEQ ID NOS: 12  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 1  
;; LENGTH: 2217  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (160)...(867)  
US-09-498-959-1

Query Match  
Best Local Similarity 1.5%; Score 52.8; DB 4; Length 2217;  
Matches 57; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 51 GGAATTGATTCCTCACTAGTACTAGGCTCGAGCGCGCGCGGCGAGGT 110  
DB 1 GAATTCGGCTTCCTCACTAGTACTAGGCTCGAGCGCGCGCGGCGAGGT 60

QY 111 CTGA 114  
DB 61 ATAA 64

RESULT 9  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-9300  
; TELEFAX: (703) 683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:

;; CLONE: pTZgpt-Fls  
US-08-232-463-14

Query Match  
Best Local Similarity 1.5%; Score 52.4; DB 1; Length 7218;  
Matches 20; Conservative 220; Mismatches 166; Indels 0; Gaps 0;

QY 2601 CCTACAAGCTGTGCCACATCTATTACAACTGGCCAGGTGTCATCTGCTTCTGCTTCTT 2660  
DB 1059 CTGCGATYY 1118

QY 2661 GCCAGTACGCCCAACAGCTGGCTTTCTGCTGGCAGAGTATTACAGAGAGCCAAATC 2720  
DB 1119 YY 1178

QY 2721 TGTCACGTCTCAACCGCTTTACTACTCTTAACCTCGAAGACGATGCGCGCTTTTC 2780  
DB 1179 YY 1238

QY 2781 TTTTGAATGACTTTGGGATTTTAAAGCTTTTATTACTTTTAACTGTTATC 2840  
DB 1239 YY 1298

QY 2841 TTCTCGATGAACCTTGGGAAGGGATAGGAGATCTAGCAATTTATTCTAGCAATGCT 2900  
DB 1299 YY 1358

QY 2901 ATTCACCGCTTCTCTTATTTATATGCTAAATAAGATTTATATTATTTATCTTCTT 2960  
DB 1359 YY 1418

QY 2961 TCTCATAGATATTTGTGAGCATTTTTTTGTTTATTATTGAAGAAAT 3006  
DB 1419 YYYYYYYYYYYYYYGTACCAATTCCTATCTCTTAACTACT 1464

RESULT 10  
US-08-487-826B-13/c  
; Sequence 13, Application US/08487826B  
; Patent No. 5993827  
; GENERAL INFORMATION:  
; APPLICANT: Sim, Kim L.  
; APPLICANT: Chitnis, Chetan  
; APPLICANT: Miller, Louis H.  
; APPLICANT: Peterson, David S.  
; APPLICANT: Su, Xin-zhaun  
; APPLICANT: Wellens, Thomas E.  
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
; AND FLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe Martens Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: California  
; COUNTRY: US  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/487,826B  
; FILING DATE: 10-SEP-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Israelsen, Ned  
; REGISTRATION NUMBER: 29,655  
; REFERENCE/DOCKET NUMBER: NIH121.001CPI  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 13:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 19124 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 US-08-487-826B-13

Query Match 1.5%; Score 51.8; DB 2; Length 19124;  
 Best Local Similarity 51.0%; Pred. No. 0.0011;  
 Matches 175; Conservative 0; Mismatches 162; Indels 6; Gaps 2;

QY 2776 TTTTCTTTTGAAGACTTTGGGATTTTAAAGCTTTTATTTTAACTTTTAACTG 2835  
 DB 15720 TTTTCTTTTGAAGACTTTGGGATTTTAAAGCTTTTATTTTAACTG 15661  
 QY 2836 TTATCTTCTGATGAAGCTTGGGAAGGGAATTAGAGATCTAGCATTTTATTTCTAGCA 2895  
 DB 15660 TTCTTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15601  
 QY 2896 TTGCTAT--TCACGGCTTCTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 2953  
 DB 15600 TTTTGTGATTAACCTTTTCAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 15541  
 QY 2954 TCTGTCTTCATAGATATTTTGTGACATTTTGTATTTTGAAGAAATGTGATA 3013  
 DB 15540 TATTTATTTTAAAAATTTTCTCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15485  
 QY 3014 AGATACCTGGAGTATTAACAGACTCTGAGAGATTTGAAAGTGTGAGATTTA 3073  
 DB 15484 AATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15425  
 QY 3074 CTTAAGTACTTTCAGAGGTGAGCAAGTCTTATTAAC 3116  
 DB 15424 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15382

RESULT 11  
 US-08-592-126-73/c  
 ; Sequence 73; Application US/08592126  
 ; Patent No. 5821091  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gregory Dolganov  
 ; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory  
 ; TITLE OF INVENTION: Polypeptides  
 ; NUMBER OF SEQUENCES: 151  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Dehlinger & Associates  
 ; STREET: 350 Cambridge Avenue, Suite 250  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94306  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/592.126  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Sholtz, Charles K.  
 ; REGISTRATION NUMBER: 38,615  
 ; REFERENCE/DOCKET NUMBER: 4600-0111  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 324-0880  
 ; TELEFAX: (415) 324-0960  
 ; INFORMATION FOR SEQ ID NO: 73:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 797 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA to mRNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 INDIVIDUAL ISOLATE: G248b.seq  
 US-08-592-126-73

Query Match 1.5%; Score 51; DB 1; Length 797;  
 Best Local Similarity 91.5%; Pred. No. 0.00027;  
 Matches 54; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 51 GGAATTCATTCATCTCAATATGACCTCATATAGGGCTGAGCGCGCCGCGGACAG 109  
 DB 797 GAATTCGCTTCATCTCAATATGACCTCATATAGGGCTGAGCGCGCCGCGGACAG 739

RESULT 12  
 US-09-141-206-1  
 ; Sequence 1; Application US/09141206  
 ; Patent No. 6187559  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Steed, Paul M.  
 ; APPLICANT: Lasala, Daniel J.  
 ; TITLE OF INVENTION: Amino Acid Sequence of Human PLD2A  
 ; FILE REFERENCE: 4-30148/P1/CGC1954/R  
 ; CURRENT APPLICATION NUMBER: US/09/141,206  
 ; CURRENT FILING DATE: 1998-08-27  
 ; EARLIER APPLICATION NUMBER: 60/057,802  
 ; EARLIER FILING DATE: 1997-08-28  
 ; NUMBER OF SEQ ID NOS: 8  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 1  
 ; LENGTH: 3388  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (162)..(2963)  
 ; OTHER INFORMATION: Human PLD2  
 ; US-09-141-206-1

Query Match 1.5%; Score 50.4; DB 4; Length 3388;  
 Best Local Similarity 98.1%; Pred. No. 0.00096;  
 Matches 51; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 62 CCATCTTAATACGACTCATATAGGGCTGAGCGCGCCGCGGACAGTCTG 113  
 DB 1 CCATCTTAATACGACTCATATAGGGCTGAGCGCGCCGCGGACAGTCTG 52

RESULT 13  
 US-08-866-757-1  
 ; Sequence 1; Application US/08866757  
 ; Patent No. 5858716  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ELISHOURBAGY, NABIL A  
 ; APPLICANT: LI, XIAOTONG  
 ; APPLICANT: BERGMA, DERK J  
 ; TITLE OF INVENTION: NOVEL 7TM RECEPTOR (H2CAA71)  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: RATNER & PRESTIA  
 ; STREET: P.O. BOX 980  
 ; CITY: VALLEY FORGE  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19482

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/866,757
  FILING DATE: 30-MAY-1997
  CLASSIFICATION: 536
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER:
  FILING DATE:
  ATTORNEY/AGENT INFORMATION:
  NAME: PRESTIA, PAUL F
  REGISTRATION NUMBER: 23,031
  REFERENCE/DOCKET NUMBER: GH-70055
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: 610-407-0700
  TELEFAX: 610-407-0701
  TELEX: 846169
  INFORMATION FOR SEQ ID NO: 1:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 4203 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: cDNA
US-08-866-757-1

Query Match
Best Local Similarity 1.4%; Score 50.2; DB 2; Length 4203;
Matches 70; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 58 GATTCATCTTAATACGACTACATATAGGCTCGAGCGCGCGCGCGGAGGTCTGAGGT 117
Db 1 GCTTCATCTTAATACACTACTATAGGCTCGAGCGCGCGCGCGGAGGTCTGAG 60

QY 118 GCAAGGACCGAGGACTAGGCGGCGGCGGCGGCTCCAGAAATAG 160
Db 61 GGAGGTGCTGTGCACCCCTCAGCAATCTGCCACCCCTACAG 103

RESULT 14
US-09-153-593-1
; Sequence 1, Application US/09153593A
; Patent No. 6174994
; GENERAL INFORMATION:
; APPLICANT: ELSHOURBAGY, NABIL A
; APPLICANT: LI, XIAOTONG
; APPLICANT: BERGSWA, DEREK J
; TITLE OF INVENTION: NOVEL 7TM RECEPTOR (H2CAA71)
; FILE REFERENCE: GH-70055-1
; CURRENT APPLICATION NUMBER: US/09/153,593A
; CURRENT FILING DATE: 1998-09-15
; EARLIER APPLICATION NUMBER: 08/866,757
; EARLIER FILING DATE: 1997-05-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 4203
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-153-593-1

Query Match
Best Local Similarity 1.4%; Score 50.2; DB 4; Length 4203;
Matches 70; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 58 GATTCATCTTAATACGACTACATATAGGCTCGAGCGCGCGCGGAGGTCTGAGGT 117
Db 1 GCTTCATCTTAATACACTACTATAGGCTCGAGCGCGCGCGGAGGTCTGAG 60

QY 118 GCAAGGACCGAGGACTAGGCGGCGGCGGCGGCTCCAGAAATAG 160
```

Db 61 GGAGGTGCTGTGCACCCCTCAGCAATCTGCCACCCCTACAG 103

## RESULT 15

US-08-998-416-595  
; Sequence 595, Application US/08998416  
; Patent No. 6239264  
; GENERAL INFORMATION:  
; APPLICANT: Philippsen, Peter  
; APPLICANT: Pohlmann, Rainer  
; APPLICANT: Steiner, Sabine  
; APPLICANT: Mohr, Christine  
; APPLICANT: Wendland, Jurgen  
; APPLICANT: Knechtle, Philipp  
; APPLICANT: Rebischung, Corinne  
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 1152  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6239264artis Corporation  
; STREET: 3054 Cornwallis Road  
; CITY: Research Triangle Park  
; STATE: No. 6239264th Carolina  
; COUNTRY: USA  
; ZIP: 27709  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/998,416  
; FILING DATE: 24-DEC-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: CH 0016/97  
; FILING DATE: 31-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meigs, J. Timothy  
; REGISTRATION NUMBER: 38,241  
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8587  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 595:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 658 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: PAG1408RP  
US-08-998-416-595

Query Match 1.4%; Score 50; DB 4; Length 658;  
Best Local Similarity 50.0%; Pred. No. 0.00045;  
Matches 151; Conservative 0; Mismatches 150; Indels 1; Gaps 1;

QY 2776 TTTTCTTTTGAATGACTTTGGGATTTTTTTAAGCTTTTACTTTTAACTG 2835  
Db 275 TATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 334  
QY 2836 TTAUCTTTCTGGATGAACCTTTGGGAGGATCTAGCATTTTATTCTAGCA 2895  
Db 335 ATTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 394  
QY 2896 TTGCTATTTCACCGCTTCTTATTTATATATGTAATAAATTAAGATTATTTATCTTC 2955  
Db 395 TTTTATTATTATGTTATTATTATTATTATTATTATTATTATTATTATTA 453  
QY 2956 TTGTTTCTCATAGATATTTTGTGAGCATTTTTTTTGTGAGAAATGCGGATAAG 3015

Db 454 TATCATAGTATAGTATATATATATATGACTTATTCATTATATAGATTGTGATTTTGTGAC 513  
QY 3016 ATACTTGGTATGATATAAAGACAGACTCTGAGAGTATTTGAATGTGTTGGAGATTACT 3075  
Db 514 ATATATATATATGCTATTTCTATTTCTATTTTATTTTATTTTATTTTATTTTATTTTATTT 573  
QY 3076 TTA 3077  
Db 574 TTA 575

Search completed: January 9, 2003, 21:37:00  
Job time : 136 secs











QY	2801	TTTTTTTAAAGCTTTTATTTACTTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTGGGA	2860
Db	1621	TTTTTTTAAAGCTTTTATTTACTTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTGGGA	1680
QY	2861	AGGGGATTTAGGAGATCTAGCATTTTATTTCTAGACATGCTATTACACGGCTTCCCTTATTT	2920
Db	1681	AGGGGATTTAGGAGATCTAGCATTTTATTTCTAGACATGCTATTACACGGCTTCCCTTATTT	1740
QY	2921	TATATGTAAGAAAATAAGATTTTATATTTTATCTTCTGTTTCTCATAGATATTTTGTGAG	2980
Db	1741	TATATGTAAGAAAATAAGATTTTATATTTTATCTTCTGTTTCTCATAGATATTTTGTGAG	1800
QY	2981	CATTTTTTTGTTTTATTTTTCGAAGAAATGGGATAAGATACATTGTTAGTATTAACAAGATC	3040
Db	1801	CATTTTTTTGTTTTATTTTTCGAAGAAATGGGATAAGATACATTGTTAGTATTAACAAGATC	1860
QY	3041	TC TGAGAGTATTTGAAATGTGTTGGAGATTTACTTTAAACGTACTTTTCAGGAGTGAGCAA	3100
Db	1861	TC TGAGAGTATTTGAAATGTGTTTCGAGATTTACTTTAAACGTACTTTTCAGGAGTGAGCAA	1920
QY	3101	GTCTCTATTATAACCTATATTAACTTTTATTTTGGAGATACCTGTTTGGATTTTAAAGGA	3160
Db	1921	GTCTCTATTATAACCTATATTAACTTTTATTTTGGAGATACCTGTTTGGATTTTAAAGGA	1980
QY	3161	GATAAGAGGCGTAAAGTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTCATATCT	3220
Db	1981	GATAAGAGGCGTAAAGTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTCATATCT	2040
QY	3221	TAAAGATATAAGGTACTATATATAAACCTATACAAAGATACAGGAGAAATATGCTTGA	3280
Db	2041	TAAAGATATAAGGTACTATATATAAACCTATACAAAGATACAGGAGAAATATGCTTGA	2100
QY	3281	TTTTTTATTTGGCAGGGGGCTAGGTGTTATCGGAGTAAAAAACAATTTGAAATTTTTTAA	3340
Db	2101	TTTTTTATTTGGCAGGGGGCTAGGTGTTATCGGAGTAAAAAACAATTTGAAATTTTTTAA	2160
QY	3341	ATTGTCCTAAAGAAAACATTTTAAGACTCTTTTAAACAAAAGCCCATGAGTAAATCTCTATA	3400
Db	2161	ATTGTCCTAAAGAAAACATTTTAAGACTCTTTTAAACAAAAGCCCATGAGTAAATCTCTATA	2220
QY	3401	TTAAACATCACTATTTATTTTGTGTTTGGACCTGGACATGATTTCTATTTGTTATATAAATAA	3460
Db	2221	TTAAACATCACTATTTATTTTGTGTTTGGACCTGGACATGATTTCTATTTGTTATATAAATAA	2280
QY	3461	AATGTGATGTCCC	3472
Db	2281	AATGTGATGTCCC	2292

US-09-873-737A-3

```

US-09-873-73/A-3
; Sequence 3, Application US/09873737A
; Patent No. US2002007697A1
; GENERAL INFORMATION:
; APPLICANT: Duke University
; APPLICANT: Lin, Haifan
; TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
; TITLE OF INVENTION: PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
; FILE REFERENCE: Attorney Docket No. US2002007697A1 180-104/2
; CURRENT APPLICATION NUMBER: US/09/873,737A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/US99/28764
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 60/110,901
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 4064
; TYPE: DNA
; ORGANISM: Mus sp.
; FEATURE:
; NAME/KEY: CDS

```

[illegible]

Db	968	TATGAGACACATCATCTCTGCAKAGAGCTGACGACAAAGTCTCCGAGGAGAACT	1027
Qy	1001	GTTTGGATTTCATGTCATCTTTATTCATCAACAGAAACATTAATTTCAAGAACAA	1060
Db	1028	GTCCTAGACCTTATGTTCAATCTATACACAGACAGAGAGACAAAGTTCCAGAGACA	1087
Qy	1061	GTTTCCAAAGAAATAATAGTTTGTGTTGTTCTTACCAAGATATACATTAAAGCATACAGA	1120
Db	1088	GTCGCAAGAGAGCTATAGAGCTCATGCTTCCACCAAGTACATTAACAAGACTACCGG	1147
Qy	1121	GTCGATGATATTGACTGGGACCCAGAAATCCCAAGACCTTTAAGAAAGCGAGCTCT	1180
Db	1148	GTCGATGACATTGACTGGGACCCAGAAATCCAAAGAGCACTTTAAGAAAGCGAGTGCCTG	1207
Qy	1181	GAGGTCAGCTTTCTTGAATATCTACAGAGAGCAATACCAACAGATTCACCACTTGAAG	1240
Db	1208	GAGGTCAGCTTTCTTGAAGTATCTACAGAGAGCAATACCAACAGAGATTCACGAGCTGAAG	1267
Qy	1241	CAGGCTGTCTTGGTCAAGCAGCCCAAGAGAGAGCGGGCCCTGAGGAGCACTCCAGGG	1300
Db	1268	CAGCCTGTCTTGGTCAAGCAGCCCAAGAGAGAGAGCGGGCCCTGAGCCTGAGC	1327
Qy	1301	CTGTCATGCTCATTCCTGAGCTCTGCTATCTTACAGTCTATCTGATTAATAGCTAT	1360
Db	1328	CCAGCTATGTCATCTCCTGAACTCTGCTATCTCACAAGCCTGATGATTAATAGCCAT	1387
Qy	1361	GATTTTAAACGTAGTAAAGACTTAAGCCCTGATCAAGAATACTCCAGACAAAGGCGAG	1420
Db	1388	GATTTTAAATGTATGAAAGACCTCGAGAGTGCACAGCGGCTGACCTCTGACGCGGAG	1447
Qy	1421	CGTGAAGTGGACGACTCATTTGATTAATTCATTCATTAACGATAATGTTCAAGGAGCTT	1480
Db	1448	CGGAGAGTGGGCGCGCTCATTCGATCAATCCACAAGAGTGCATATGTCAGAGAGACTT	1507
Qy	1481	CGAGCTGGGGTGTGAGCTTGTGATTCACATTAATCTGCTCTTCTCAGAGAAATTTTGAA	1540
Db	1508	CGAGCTGGGGCTGAGCTTGCATTAATCTTGTGCTGCTCTCTCGAGAAATCTTAAACA	1567
Qy	1541	ACAGAAAGATTTCAACAAAGTGAAGAAAACATTTGATTACATCCACAATTTGACAGTTGG	1600
Db	1568	TCGGAAGATTCACACAGGCGGAGAAAGAGTTTGATTCACACCAATTTGACAGACTGG	1627
Qy	1601	TCGAAAGAAACAAGAGGTGACCATTAATTAAGGTTAAGCCACTAGATTAATCTGCTGTG	1660
Db	1628	TCGAAAGAAACAAGAGGCGCGCGCTGATCAGGTGAGAACCATTTGAAATCTGCTCTG	1687
Qy	1661	ATCTATACGCGAAGAAATTATGAGACGCCAATTCATTTGATTAATCTATTTAAAGTT	1720
Db	1688	ATCTATACCGGAGGAAATTATGAGACGCCAATCTCATGATTAACAACTGTTCAAAGTG	1747
Qy	1721	ACACCAAGCATGGGATTCGAAATGAGAAAGCAATTAAGTAAAGTGGATGACAGAACT	1780
Db	1748	ACTCCAGCATGGGATTCAGATTAATAAAGCAATATATGAGAGTGGATGACAGAAACA	1807
Qy	1781	GAGGCTTATTAAGTCTTACAGCAAAAGTCAAGACACCAAGATGTTGTCTGT	1840
Db	1808	GAGGCTTATCTGAGAGCTTCAAGCAAGAGTACGTCAGACATCAATATGTTGTCTGT	1867
Qy	1841	CTGTTGTAAGTATTCGGAAGACAAATAGATGCTATTAATAAATACTGTGTAAGAT	1900
Db	1868	CTCTTGTCAAGTATTCGGAAGACAAATATGATGCTATCAAGATGATCTTGTGTACAGAC	1927
Qy	1901	TGCGCTTACCCCAAGTCAGTGTGGTGGCGGAACTTATAGCAACAGCAACTGTCTATG	1960
Db	1928	TGCGCTTACCCCAAGTCAGTGTGGTGGCGGAACTTATAGCAACAGCAACTGTCTATG	1987
Qy	1961	GCAATGCTTACAAAGTTGCCCTACAGATGAATCTGCAAGAATGGGAGAGAGCTCTGAGAG	2020
Db	1988	GCAATGCTTACCAAGATTCGCCCTCGAGTGAATCTGCAAGATGGGAGAGAGCTCTGAGAG	2047
Qy	2021	GTCGATATCCCTCGAAGCTCTGTATGATCTGTTGGCATTCATGTTTACATGACATGACA	2080
Db	2048	GTCGATATGCGCTCTGAATCTGCGCATGATCTGTGGGATTCAGCTTACATGACATGACACA	2107

Qy	2081	GCTGGCGGAGGTCAATCGCAGGATTTGTTCGCGAGCATCAATGAAGGATGACCGCGTGG	2140
Db	2108	GCTGGCGGAGGTCCATCGCAGGATTCGTCGCGAGCATCAATGAAGGATGACCGCGTGG	2167
Qy	2141	TTCTCACCGCTGCATATTTCCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAAGTCGC	2200
Db	2168	TTCTCCCCGCTCGCTTTCCAGGACCGCGCGAGGAGCTGGTGGATGTTCTCAAGTGTGC	2227
Qy	2201	CTGCAAGCGGCTCTGAGGGCTTGGAAATAGCTGCAATGAGTACATCCCGACGCGGATCATC	2260
Db	2228	TTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCGCGCTGTCTATC	2287
Qy	2261	GTGTACCGGATGCGGTAGAGACGGCGAGCTGTAAGAACACTGGTGAATACGAAAGTGCCA	2320
Db	2288	GTGTACCGAGACGGTGTGGGGGACGGCGAGCTGAAGACCTCGTCAATTATGAGGTCCCA	2347
Qy	2321	CAGTTTTTGGATTTGTCTAAAATCCATTGGTGTAGAGTTTACAACCTCAGANTAAACGGTAAT	2380
Db	2348	CAGTTCCTAGATTGCCTCAAGTCAAGTCAGTCTGGGAGAGGTTTAAACCCAGACTGACTGTAATC	2407
Qy	2381	GTGTGAAGAAAGAGTGAACACCAAGATTTTTTGTCTAGTCTGGAGGAAGACTTCAGAAT	2440
Db	2408	GTGTGAAGAAAGCGTGTCAATGCCAGGTTTTTTGTCTAGTCTGGGGGAAGACTTCAGAAC	2467
Qy	2441	CCACTTCCTGGAAACAGTTATTGATGTAGAGGTTACCGACAGAACTGGTAGATGATTTTTT	2500
Db	2468	CCTCTTCAGGGACAGTCAATCGATGTGAAGTTCACGACAGAGTGGTAGATGATTTTTTC	2527
Qy	2501	ATCGTGAGCCAGGCTGTGAGAAAGTGGTAGTGTTCCTCCACACATTTACAATGTCATCTAT	2560
Db	2528	ATCGTGAGCCAGGACGTGAGAAAGCGGAGTGTGTCCCAACACACTACAATGTGATCTAT	2587
Qy	2561	GACAAACGCGCTGAAGCCGACACACACATACAGCGCTTGACCTACAAGCTGTGCCACATC	2620
Db	2588	GACAGCAGTGGCTTGAAGCCCGACACATCCAGCGGCTGACATACAAGNTCTGTCCACGTCG	2647
Qy	2621	TATTACAACCTGGCCAGGTGTCAATTCGTCTCTCTCTTGGCAGTAGCGCCACCAAGCTG	2680
Db	2648	TACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGGCCAGTAGTCACACAAGCTG	2707
Qy	2681	GCTTTTCTTTTGGCCAGAGTATTCACAGAGAGCAATCTGTCTACTGTCTAAACCGCTTT	2740
Db	2708	GCCTTCTCTGTGGCCAGAGCATCCACAGAGGCAAAACCTCTCCCTGTCCAAACCGCTC	2767
Qy	2741	TACTACTCTTAACCTGCAG	2759
Db	2768	TACTACTCTTAACCAAGCAG	2786

```

RESULT 4
US-09-998-598-2600/c
; Sequence 2600, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Megghet, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 2600
; LENGTH: 582
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 527..531,553

```

OTHER INFORMATION: n = A, T, C or G				
US-09-998-598-2600				
Query Match	15.9%	Score 552.2;	DB 10;	Length 582;
Best Local Similarity	99.3%	Pred. No. 2.8e-115;		
Matches 551; Conservative	3;	Mismatches 1;	Indels 0;	Gaps
QY 2915	TTATTTTATATGTAAATAAATTAAGATTTTATATTTTATATCTCTGTTCCTCATAGATATTT	2974		
Db	582 TTATTTTATATGTAAATAAATTAAGATTTTATATTTTATATCTCTGTTCCTCAWAGAAATTT	523		
QY 2975	TCTGAGCAATTTTTTGTTTTATTTTGAAGAAATGTGGATAAGATATCTCGTAGTATAAAC	3034		
Db	522 TCTGAGCAATTTTTTGTTTTATTTTGAAGAAATGTGGATAAGATATCTCGTAGTATAAAC	463		
QY 3035	AGACCTCTGAGAGATTTTGAATGTGTTGGAGATTTACTTTAAACGTACTTTCCAGGAGT	3094		
Db	462 AGACTCTCTGAGAGATTTTGAATGTGTTGGAGATTTACTTTAAACGTACTTTCCAGGAGT	403		
QY 3095	GAGCAAGTCCTACTTATAAACCTTATATACTTTTATTTTGGAGATACCTGTTTGGATTT	3154		
Db	402 GAGCAAGTCCTACTTATAAACCTTATATACTTTTATTTTGGAGATACCTGTTTGGATTT	343		
QY 3155	AAAGGAGATAGAGCGGTAAAGTAGATGCTCACTAACCATAGGTGGGTTCAGCTC	3214		
Db	342 AAAGGAGATAGAGCGGTAAAGTAGATGCTCACTAACCATAGGTGGGTTCAGCTC	283		
QY 3215	ATATCTTTAAAGATAAAAGGTACTTATTATATACCTATACACAGATACAGGAGAAATAT	3274		
Db	282 ATATCTTTAAAGATAAAAGGTACTTATTATATAACCTATACACAGATACAGGAGAAATAT	223		
QY 3275	GCTTGATTTTTTATTTGGCAGGGGGCTAGTTGTATGGAGTAAAAAAAACATTGAAAAT	3334		
Db	222 GCTTGATTTTTTATTTGGCAGGGGGCTAGTTGTATGGAGTAAAAAAAACATTGAAAAT	163		
QY 3335	TTTTAAATTTGTCAAAAGAACATTTTAAAGATCTTTTAAACAAAAAGCCATGAGTAAATC	3394		
Db	162 TTTTTAAATTTGTCAAAAGAACATTTTAAAGATCTTTTAAACAAAAAGCCATGAGTAAATC	103		
QY 3395	TCTATATTAAACATCACTATTTATTTTGTTTTGGAACTGGGACATGATTTCTATTTGTTATA	3454		
Db	102 TCTATATTAAACATCACTATTTATTTTGTTTTGGAACTGGGACATGATTTCTATTTGTTATA	43		
QY 3455	AAATAAAATTTGATG 3469			
Db	42 AAATAAAATTTGATG 28			

```

RESULT 5
US-09-919-580-634/c
; Sequence 634, Application US/09919580
; Patent No. US20020110832A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF COLON CANCER
; FILE REFERENCE: 210121.552
; CURRENT APPLICATION NUMBER: US/09/919,580
; CURRENT FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 634
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 535, 539, 561
; OTHER INFORMATION: n = A,T,C or G
US-09-919-580-634

```

Query Match 15.9%; Score 550.4; DB 10; Length 590;  
Best Local Similarity 99.3%; Pred. No. 7.2e-115;  
Matches 551; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2915 TTATTTATATGTAATAATAGATTTATATTTATCTCTGTTCTCATAGATTTT 2974  
DB 590 TTATTTATATGTAATAATAGATTTATATTTATCTCTGTTCTCATAGATTTT 531  
QY 2975 TGTGAGCAATTTTGTGTTATTTTGAAGAAATGCGATTAAGTCTGTAGATTAAC 3034  
DB 530 TGTGAGCAATTTTGTGTTATTTTGAAGAAATGCGATTAAGTCTGTAGATTAAC 471  
QY 3035 AGACTCTCTGAGGATTTGGAATGTGTTGAGATTTACTTAACGTAATCTTCAAGAGT 3094  
DB 470 AGACTCTCTGAGGATTTGGAATGTGTTGAGATTTACTTAACGTAATCTTCAAGAGT 411  
QY 3095 GAGCAATCTCTACTTAATTAACCTATATTTATTTTGAATGCTGTTTGAATTT 3154  
DB 410 GAGCAATCTCTACTTAATTAACCTATATTTATTTTGAATGCTGTTTGAATTT 351  
QY 3155 AAAGAGATTAAGGCGTAAAGTAGATGCTCACTAACCATAGGTGGGTTTCAGCTC 3214  
DB 350 AAAGAGATTAAGGCGTAAAGTAGATGCTCACTAACCATAGGTGGGTTTCAGCTC 291  
QY 3215 ATATCTTAAGATTAAGGCTATATATATTAACCTATTAACAGATTAAGGAAATAT 3274  
DB 290 ATATCTTAAGATTAAGGCTATATATATTAACCTATTAACAGATTAAGGAAATAT 231  
QY 3275 GCTTGAATTTTATTTGAGGAGGCGTATGTTGATGAGGATTAAGGAAATAT 3334  
DB 230 GCTTGAATTTTATTTGAGGAGGCGTATGTTGATGAGGATTAAGGAAATAT 171  
QY 3335 TTTTAAATTTGTCAAAGAAACATTTTAAAGCTCTTTAAACAAAAGGCCATGATTAATC 3394  
DB 170 TTTTAAATTTGTCAAAGAAACATTTTAAAGCTCTTTAAACAAAAGGCCATGATTAATC 111  
QY 3395 TCTATATTAACATCACTATTTATTTGTTGGAATGGAATGATTTATTTGTTATA 3454  
DB 110 TCTATATTAACATCACTATTTATTTGTTGGAATGGAATGATTTATTTGTTATA 51  
QY 3455 AAATAAATTTGATGT 3469  
DB 50 AAATAAATTTGATGT 36

RESULT 6  
US-09-920-300A-34  
Sequence 34, Application US/09920300A  
Patent No. US20020136728A1  
GENERAL INFORMATION:  
APPLICANT: King, Gordon E.  
APPLICANT: Meagher, Madeleine Joy  
APPLICANT: Xu, Jiangchun  
APPLICANT: Secretist, Heather  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
FILE REFERENCE: 210121.547  
CURRENT APPLICATION NUMBER: US/09/920.300A  
CURRENT FILING DATE: 2001-07-31  
NUMBER OF SEQ ID NOS: 1789  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 34  
LENGTH: 501  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-920-300A-34

Query Match 13.2%; Score 459.4; DB 10; Length 501;  
Best Local Similarity 99.8%; Pred. No. 2.3e-94;  
Matches 460; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2695 CCAGAGTATTCACAGAGCAAAATCTGTCACTGCAACCGCTTTACTACTTAACC 2754  
DB 121 TATTTACTTTTATTTTACCTGTTATCTTCTGATGAAATCTTGGAGATTTAGGAGA 2874  
QY 2875 TCTAGCAATTTATTTCTAGCATTTGATTAACCGGCTCTTATTTATGTAATAATTT 2934  
DB 181 TCTAGCAATTTATTTCTAGCATTTGATTAACCGGCTCTTATTTATGTAATAATTT 240

DB 1 CCAGAGTATTCACAGAGCAAAATCTGTCACTGCAACCGCTTTACTACTTAACC 60  
QY 2755 TCCAGAAAGCATGAGAGCGGCTTTCTTTTGAATGACTTTGGAGTTTATTAAGCTTT 2814  
DB 61 TCCAGAAAGCATGAGAGCGGCTTTCTTTTGAATGACTTTGGAGTTTATTAAGCTTT 120  
QY 2815 TATTTACTTTTATTTTACCTGTTATCTTCTGATGAAATCTTGGAGATTTAGGAGA 2874  
DB 121 TATTTACTTTTATTTTACCTGTTATCTTCTGATGAAATCTTGGAGATTTAGGAGA 180  
QY 2875 TCTAGCAATTTATTTCTAGCATTTGATTAACCGGCTCTTATTTATGTAATAATTT 2934  
DB 181 TCTAGCAATTTATTTCTAGCATTTGATTAACCGGCTCTTATTTATGTAATAATTT 240  
QY 2935 AAGATTTATTTATTTATCTTCTGTTCTCAATGATTTTGGAGATTTTGTGTTA 2994  
DB 241 AAGATTTATTTATTTATCTTCTGTTCTCAATGATTTTGGAGATTTTGTGTTA 300  
QY 2995 TTTTGAAGAAATGAGATTAAGTCTGTAGTATTAAGAAAGCACTCTGAGAGATTTTG 3054  
DB 301 TTTTGAAGAAATGAGATTAAGTCTGTAGTATTAAGAAAGCACTCTGAGAGATTTTG 360  
QY 3055 AAATGCTTTGAGATTTACTTAACGTAATCTTTCAGAGTGAAGCAATCTTACTATTA 3114  
DB 361 AAATGCTTTGAGATTTACTTAACGTAATCTTTCAGAGTGAAGCAATCTTACTATTA 420  
QY 3115 CCTATATTAATTTATTTTGAATGATACCTGTTTGAATTTA 3155  
DB 421 CCTATATTAATTTATTTTGAATGATACCTGTTTGAATTTA 461

RESULT 7  
US-10-033-528-34  
Sequence 34, Application US/10033528  
Patent No. US20020131971A1  
GENERAL INFORMATION:  
APPLICANT: King, Gordon E.  
APPLICANT: Meagher, Madeleine Joy  
APPLICANT: Xu, Jiangchun  
APPLICANT: Secretist, Heather  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
FILE REFERENCE: 210121.547C1  
CURRENT APPLICATION NUMBER: US/10/033.528  
CURRENT FILING DATE: 2001-12-26  
NUMBER OF SEQ ID NOS: 1896  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 34  
LENGTH: 501  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-033-528-34

Query Match 13.2%; Score 459.4; DB 12; Length 501;  
Best Local Similarity 99.8%; Pred. No. 2.3e-94;  
Matches 460; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2695 CCAGAGTATTCACAGAGCAAAATCTGTCACTGCAACCGCTTTACTACTTAACC 2754  
DB 1 CCAGAGTATTCACAGAGCAAAATCTGTCACTGCAACCGCTTTACTACTTAACC 60  
QY 2755 TCCAGAAAGCATGAGAGCGGCTTTCTTTTGAATGACTTTGGAGATTTTAAAGCTTT 2814  
DB 61 TCCAGAAAGCATGAGAGCGGCTTTCTTTTGAATGACTTTGGAGATTTTAAAGCTTT 120  
QY 2815 TATTTACTTTTATTTTACCTGTTATCTTCTGATGAAATCTTGGAGATTTAGGAGA 2874  
DB 121 TATTTACTTTTATTTTACCTGTTATCTTCTGATGAAATCTTGGAGATTTAGGAGA 180  
QY 2875 TCTAGCAATTTATTTCTAGCATTTGATTAACCGGCTCTTATTTATGTAATAATTT 2934  
DB 181 TCTAGCAATTTATTTCTAGCATTTGATTAACCGGCTCTTATTTATGTAATAATTT 240



QY 2935 AAGATTTTATATTTTATCTTCTGTTTCTCATAGATATTTTGTGAGCATTTTGTGTTT 2994  
 DB 241 AAGATTTTATATTTTATCTTCTGTTTCTCATAGATATTTTGTGAGCATTTTGTGTTT 300  
 QY 2995 TTTTGAAGAAATGTTGATAAGATATCTGTTAGTATATAAACAGACTCTCTGAGATTTTG 3054  
 DB 301 TTTTGAAGAAATGTTGATAAGATATCTGTTAGTATATAAACAGACTCTCTGAGATTTTG 360  
 QY 3055 AAATGTTTGGAGATTTTACTTAAACGTTACTTTCAGGAGTGAAGCTCTTACTTTATAAA 3114  
 DB 361 AAATGTTTGGAGATTTTACTTAAACGTTACTTTCAGGAGTGAAGCTCTTACTTTATAAA 420  
 QY 3115 CCTATATTAACCTTTATTTTGTAGATACCTGTTTGTGAATTTA 3155  
 DB 421 CCTATATTAACCTTTATTTTGTAGATACCTGTTTGTGAATTTA 461

RESULT 8  
 US-09-777-564-1706/c  
 ; Sequence 1706, Application US/09777564  
 ; Patent No. US20020022591A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Algate, Paul A.  
 ; APPLICANT: Mannion, Jane  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
 ; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER  
 ; FILE REFERENCE: 210121.493  
 ; CURRENT APPLICATION NUMBER: US/09/777,564  
 ; PRIOR FILING DATE: 2001-02-05  
 ; NUMBER OF SEQ ID NOS: 1730  
 ; SOFTWARE: FastSeq for Window Version 4.0  
 ; SEQ ID NO 1706  
 ; LENGTH: 460  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-09-777-564-1706

Query Match 13.2%; Score 458.4; DB 10; Length 460;  
 Best Local Similarity 99.8%; Pred. No. 3.7e-94;  
 Matches 459; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2695 CCAGAGTATTCACAGAGAGCCAAATCTGTCACCTGTCACAAACCGCTTTTACTTACTCTTAACC 2754  
 DB 460 CCAGAGTATTCACAGAGAGCCAAATCTGTCACCTGTCACAAACCGCTTTTACTTACTCTTAACC 401  
 QY 2755 TGCAAGAGAGCATGCAGCGCTTTTCTTTTGTAAATGACCTTTGGGATTTTAAAGCTTT 2814  
 DB 400 TGCAAGAGAGCATGCAGCGCTTTTCTTTTGTAAATGACCTTTGGGATTTTAAAGCTTT 341  
 QY 2815 TATTTACTTTTTTTTAACTGTTATCTTCTTCTGATGAACTTTGGGATTTTAAAGCTTT 2874  
 DB 340 TATTTACTTTTTTTTAACTGTTATCTTCTTCTGATGAACTTTGGGATTTTAAAGCTTT 291  
 QY 2875 TCTAGCATTTTATTTCTAGCATTTCTATTCACCGCTTCTCTTATTTATATGTAATAAAT 2934  
 DB 280 TCTAGCATTTTATTTCTAGCATTTCTATTCACCGCTTCTCTTATTTATAGTAAAT 221  
 QY 2935 AAGATTTTATATTTTATCTTCTGTTTCTCATAGATATTTTGTGAGATTTTGTGTTT 2994  
 DB 220 AAGATTTTATATTTTATCTTCTGTTTCTCATAGATATTTTGTGAGATTTTGTGTTT 161  
 QY 2995 TTTTGAAGAAATGTTGATAAGATATCTGTTAGTATATAAACAGACTCTCTGAGATTTTG 3054  
 DB 160 TTTTGAAGAAATGTTGATAAGATATCTGTTAGTATATAAACAGACTCTCTGAGATTTTG 101  
 QY 3055 AAATGTTTGGAGATTTTACTTAAACGTTACTTTCAGGAGTGAAGCTCTTACTTTATAAA 3114  
 DB 100 AAATGTTTGGAGATTTTACTTAAACGTTACTTTCAGGAGTGAAGCTCTTACTTTATAAA 41  
 QY 3115 CCTATATTAACCTTTATTTTGTAGATACCTGTTTGTGAATTT 3154  
 DB 40 CCTATATTAACCTTTATTTTGTAGATACCTGTTTGTGAATTT 1

RESULT 9  
 US-09-801-574-15  
 ; Sequence 15, Application US/09801574  
 ; Patent No. US20020081592A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Peijing Jeremy  
 ; APPLICANT: Page, David C.  
 ; TITLE OF INVENTION: Reproduction-Specific Genes  
 ; FILE REFERENCE: 0399.2007-002  
 ; CURRENT APPLICATION NUMBER: US/09/801,574  
 ; CURRENT FILING DATE: 2001-03-07  
 ; PRIOR APPLICATION NUMBER: 60/187,518  
 ; PRIOR FILING DATE: 2000-03-07  
 ; PRIOR APPLICATION NUMBER: 60/261,557  
 ; PRIOR FILING DATE: 2001-01-12  
 ; NUMBER OF SEQ ID NOS: 90  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 15  
 ; LENGTH: 3649  
 ; TYPE: DNA  
 ; ORGANISM: Mus musculus  
 ; US-09-801-574-15

Query Match 10.9%; Score 377; DB 10; Length 3649;  
 Best Local Similarity 52.5%; Pred. No. 2.2e-75;  
 Matches 974; Conservative 0; Mismatches 862; Indels 21; Gaps 6;  
 QY 914 CCTGGCTTCACTACTTCCATCTCTTCAGTATGAAAAACAGCATCATGCTCTGCACCTGAGTT 973  
 DB 1 CCTGGCTATGCGGTAGTATCCGAGGACAGACGGGGTCTCTTCTCTGCTCGCTGATGC 60  
 QY 974 AGCCATAAAGTCTTCCAGTGAGACTGTTTGTGATTTTCACTTTTATCATCAG 1033  
 DB 61 TCTATAAGGTTCATTCGGAACGACTCTGTGCTGATGCTCAIGCATGCTATCTACCCAGCAG 120  
 QY 1034 ACAGAAGAACATAAATTTTCAAGAACAAAGTTTCCAAAGAAATAATAGTGTGTGTTTCTT 1093  
 DB 121 AACAAAGAGCAC---TTCCAGGACGAGTGCAGCAAGCTTCTGGTTGGCAGCATTTGTCTATC 177  
 QY 1094 ACCAAGTATAACATTAAGACATACAGAGTGGATGATATTGACTGGGACCAAGATCCCAAG 1153  
 DB 178 ACGGCTACAAACAATCGTACCTCCGAATCGATGATGTGGACTGGAAACAAAGACCCCTAAA 237  
 QY 1154 AGCACTTTTAAAGAAAGCGCGCTCTGGGCTCAGCTTCTTTAGAAATACTACAGGAAGCAA 1213  
 DB 238 GACAGCTTTTGTCTGTCGAGCGGAAGAAATACATTTCTTGGATATCTACAGCAAAAC 297  
 QY 1214 TACAACCAAGAGATCAACGACTTGAAGCAGCTCTGTTGGTCAAGCCAGCCCA---AGAGA 1270  
 DB 298 TATGGGATCACAGTCAAGGAAGATGACCAAGCGCTGTGATCCACCGCCCGCAGTGAGAGA 357  
 QY 1271 AGCGGGGCTTGGGGGACACTGCCAGGCGCTGCCATGCTCATTCTGAGCTCTGCTAT 1330  
 DB 358 CAGATAAACATGCGATGTTGCTGAAGGCGAGATCTCTGCTGCTGCCGAGCTCTCTTC 417  
 QY 1331 CTTCACAGTCTAACTGATAAATAATGCGTAATGATTTTAAACGATGAAGAGCTTTAGCCGTT 1390  
 DB 418 ATGACGGGATCCCTGAGAGATGAAGAGGACTTTCAGGCGCATGAAGGACTTGAAGCTCAG 477  
 QY 1391 CATACAGACTTAATCTCAGAGCAAGGCGTGAAGTGGGACGACTCATTTGATTACATT 1450  
 DB 478 CAGATTAACCTTGAGCCCCCAAGCAGCACCCCGTGTGTTTGGAAATGCTCTCTCAGAGAAAT 537  
 QY 1451 CATAAAAACGATAATGTTTCAAGGGGAGCTTCGAGACTGGGGTTTGGAGCTTTGATTCCAAAC 1510  
 DB 538 TCACAAAACGACAGCAGCAATGAGCTGCCCGCTGGGGCTCAGTCTGCATAAAGAT 597  
 QY 1511 TTACTGTCTCTCAGGAAGAAATTTTGCAAAACAGAAAAGATTCCACAGGTGGAAAAACA 1570  
 DB 598 GTCCACAGAGTTGAAGGTGCGGCTTCTGCAATGGAGAGGATCAACTTAAGGAACACTTCA 657  
 QY 1571 TTTGATTACATCCACAAATTTTCAGATTTGGTCCAAAGAAAAACAGAGGTGCACCAATTAAT 1630

Db 658 TTGTG---CACATCGAGGCGCTGAACCTGGTTAAGAGTGAACAGAGATGCTTCATT 714  
 Qy 1631 AGGTGAAGCCCTGATGATACTGGCTGTGATCTATACCGAAGAAATTATGAGCAGCC 1690  
 Db 715 CTACTATTCCATGATCTTCTGGGCACTTTTATCCAAAGAGAGCAATGACCAAGCC 774  
 Qy 1691 AATTCAATGATACAAATCTATTATTAAGTTACACAGCCTAGCGGCAATGAGAAA 1750  
 Db 775 AGAGAACTGGTTAACTGTTGAGAAAGATTGCGGGCCATGGGATGCGCAAGCCCC 834  
 Qy 1751 GCATATATGTTAAAGT---GATGACAGAACTGAAGCTTAAAGTCTTAACAGCA 1807  
 Db 835 CAGGCTGGGTGAGCTGAAGATGACCGAATAGAGAACTTATAGAGCACTTACAGTCC 894  
 Qy 1808 AAGGTACAGCAGAC-----ACCCAGATAGTTGTCTGTCTGTGTGCAAGTAAAGTGA 1861  
 Db 895 TTACTGGAGTTGAGGGAGATACAAATGATGCTTGTGATCATCATGAGGACACGTGAT 954  
 Qy 1862 GACAATATGATGCTATTAAATAATCTGTGTATACAGATTCCTTACCCCAAGTGT 1921  
 Db 955 GATCTCATGAGGACATCAAGAGCTGTGTGTGTGCAAGTCCCAAGTGTCTCAGAGTTC 1014  
 Qy 1922 GTGTGCGCCGACCTTGAAGCAACAGCAAACTGTATGCTGATTCCTTCAAAAGTGTCC 1981  
 Db 1015 ATCATATGCGAATCATTTGGTCAGCCACAGGCTTCTGAGCGTGTGCTAGAAAATTTTA 1074  
 Qy 1982 CTACAGATGAATCTGCAAGATGAGGAGAGAGCTGTGAGGGTGAACATCCCTGAAAGTTC 2041  
 Db 1075 CTTCAGATGAATCTGAACTGAGGAGTGTGAGTCTTGGGAGATGATATTCCTGAAAGCA 1134  
 Qy 2042 GTGATGATCGTTGAGATGATTTGTTACATGATGATGACAGCTGGGCGAGAGTCAATCGCA 2101  
 Db 1135 CTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1194  
 Qy 2102 GGATTTGTTGCTGAGATCAATGAAGAGATGACCGCTGGTCTTCAAGCTGATATTTTCA 2161  
 Db 1195 GGTCTGTGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1254  
 Qy 2162 GATGAGGAG 2221  
 Db 1255 ATGCGCATCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1314  
 Qy 2222 TGGAAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2281  
 Db 1315 TATTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1374  
 Qy 2282 GACGGGAG 2341  
 Db 1375 GATGGGAG 1434  
 Qy 2342 TCCATTTGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2401  
 Db 1435 GCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1491  
 Qy 2402 ACCAGATTTTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2461  
 Db 1492 ACCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1551  
 Qy 2462 GATGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2521  
 Db 1552 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1611  
 Qy 2522 AGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2581  
 Db 1612 CAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1671  
 Qy 2582 GACCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2641  
 Db 1672 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1731  
 Qy 2642 ATTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2701

Db 1732 ATCCGAGTTCCAGCTCTGTGAGATATGCCACAGAGTAGCTTTCTGTCCGAGAGATT 1791  
 Qy 2702 ATTCAAGAGAGCCAAATCTGTACTGTCAACCGCCTTACTACCTTAACTGCA 2758  
 Db 1792 TTGATCATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1848  
 RESULT 10  
 US-09-873-737A-1  
 Sequence 1, Application US/09873737A  
 Patent No. US20020076797A1  
 GENERAL INFORMATION:  
 APPLICANT: Duke University  
 APPLICANT: Lin, Haifan  
 TITLE OF INVENTION: PURIFIED AND ISOLATED p1w1 FAMILY GENES AND GENE  
 TITLE OF INVENTION: PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME  
 FILE REFERENCE: Attorney Docket No. US20020076797A1 180-104/2  
 CURRENT FILING DATE: 2001-06-04  
 PRIOR FILING DATE: 1999-12-03  
 PRIOR APPLICATION NUMBER: PCT/US99/28764  
 PRIOR FILING DATE: 1998-12-04  
 NUMBER OF SEQ ID NOS: 21  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 1  
 LENGTH: 3047  
 TYPE: DNA  
 ORGANISM: Drosophila sp.  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (84)..(2612)  
 NAME/KEY: misc feature  
 LOCATION: (120)  
 OTHER INFORMATION: n=a or c, Xaa=Leu or Ile  
 NAME/KEY: misc feature  
 LOCATION: (399)  
 OTHER INFORMATION: n=a or t, Xaa=Leu or Ile  
 NAME/KEY: misc feature  
 LOCATION: (2436)  
 OTHER INFORMATION: n=a or c, Xaa=Leu or Ile  
 US-09-873-737A-1  
 Query Match 7.3%; Score 253; DB 10; Length 3047;  
 Best Local Similarity 47.1%; Pred. No. 1.9e-47;  
 Matches 1080; Conservative 0; Mismatches 1184; Indels 29; Gaps 9;  
 Qy 470 GATCAAAAAGAGTTCTTCAAGCATTAATGTAAGTTAAGCTTAACCACTTCCGGCTG 529  
 Db 342 GATCCAG 401  
 Qy 530 AATCCCGTCCCGAGTGGGCTTATATGATGATGATGATGATGATGATGATGATGATGATGATG 589  
 Db 402 AAAACCAAGCCGAAATGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 461  
 Qy 590 GCCAG 649  
 Db 462 AATCCCGTCCCGAGTGGGCTTATATGATGATGATGATGATGATGATGATGATGATGATGATG 521  
 Qy 650 GCTTTGATGAGAGATTAATTTTAACTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 709  
 Db 522 CTATTCGAG 581  
 Qy 710 TTATGTAAG 769  
 Db 582 AGCGAAG 641  
 Qy 770 CACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 829  
 Db 642 TCGTGTGCTGAGCCCGCTTTTGGCAAGTCTTAATTAATTTGGCGGCTGATGAAG 701  
 Qy 830 ATATGATTTGCAAGAAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 889

Db 702 GGCCTAAATTTGGAAATTTAGTTGGCGGTAATCTCTTTGATCCCGAGCTAAGATCGAAATA 761  
 Qy 890 CCAAGTACAGGTTGGTGAATTTGGCTGGCTGCTCACTTCCATCTCTTCAAGTATGAAAC 949  
 Db 762 AGGAGTTCAAAATGGAGCTATGGCCGGCTATGAGACATCGATTCGTGACGACGAAGA 821  
 Qy 950 AGCATCATGCTCTGCACTGAGCTTAGCCATAAAGTCTTCCGAGTACAGACTGTTTGGAT 1009  
 Db 822 GATATTTTATTTGGGCAACGAATAAATCTCAAAAGTTATGGCCAGGAGCATCTACGAC 881  
 Qy 1010 TTCAATGTTCAATTTTATCATCAGACAGAAAGACATAAATTTTCAAGAAACAAGTTTCCAAA 1069  
 Db 882 ATAATGGAGCTTGTCTCACAAATCCGGCTCGTCAATCAGGACGAGTACGGT-----A 935  
 Qy 1070 GAATAATAGCTTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGAGTGAATGAT 1129  
 Db 936 AATGTTTGGACTTGTATGTTCTTACGGATTAACAATAACAGAACTTATCGTATCAATGAT 995  
 Qy 1130 ATTGACTGGACCAAGATCCCAAGAGCACCTTTAAGAAAGCCGAGCTCTGGGTGACG 1189  
 Db 996 GTGACTTTGGACAAATCTCGGAATCAACATTT---CAGTTGCAAGGTAGAGATATCAGT 1052  
 Qy 1190 TTCTTAGAATACAGGAGGACATACCAAGAGATCACCGACTTGAAGCAGCTGTC 1249  
 Db 1053 TTGTTGATGATCTCTCACTAAATATAATATACGCAATTCGCGACCAACAATCAGCGCTG 1112  
 Qy 1250 TTGGTCAGCCAGCCCAAGAGAGGGGGCCCTGGGGGACACTGCCAGGSCCTGCCATG 1309  
 Db 1113 CTGATTTTCCAAA---AATAGGACAGGCTCTAATAAATCAGCTAGGGAATAGTGTA 1169  
 Qy 1310 CTATCTCTGAGCTCTGCTATCTTACAGTCTAATGATTAATAAATCGCTAATTTTAAAC 1369  
 Db 1170 CTAAATCTGAGCTCTGCGAGTGTGCTGAGTCAATGCGAGATGCGCTCAAACTTTCAG 1229  
 Qy 1370 GTGATGAAGACTTACCGCTTCTATACAGACTTACTCCAGAGCAAGGAGCGTGAAGTG 1429  
 Db 1230 CTATGCGTGCATGAGCAGTATACGCAATGAAACCCCAACCAACCACTGATCGATTG 1289  
 Qy 1430 GGACACTCATTTGATTAATCATTAATAAAGATTAATGTTTCAAGGGAGCTTCCAGACTGG 1489  
 Db 1290 CGCGCTTTTAAACACCGTTTCAAAACACTCCAGAAAGTGTGAAGTCTTGAAGACTGG 1349  
 Qy 1490 GGTGAGCTTGTGTTTCAACTCTGCTTCTTCCAGAGAAATTTGCAAAAGAGAAAG 1549  
 Db 1350 AACATGGAATCTGACAGAACGTCACAGAGTACAGGCGCGGATTAATGACAGCAGAAC 1409  
 Qy 1550 AT---TCACCAAGTGGAAAAACAATTTGATTAACAATCCACAATTTGACAGTTGGTCCAA 1605  
 Db 1410 ATCGTGTTCATAATGGAAGGTTCTGCTGGAGAAACCGCTGATTTGGCAAGGCACTTC 1469  
 Qy 1606 AGAAACAGAGTGCACCAATTAATAGTGTAAAGCCACTAGATAAATCTGCTGTGATCTTA 1665  
 Db 1470 AGAGCCAAAGGATGTTTACCACTCCAGAGGATGGC-CTCGATCGTGGGCTGTATCGC 1528  
 Qy 1666 TACGCGAAGAAATPATGAAGCAGCAATTCATTTGATACAAAATCTATTTAAAGTTTACCC 1725  
 Db 1529 GCCGCAAGGAAATCCCACTGAACTCCGAATCTTACTTGTACTTTGTATGAGCAGCTAG 1588  
 Qy 1726 AGCCATGGCATGCAATGAGAAAGC---ATAATGATGGAAGTGGATGACAGAACTGA 1782  
 Db 1589 TGGATGGGTCTTAGAATTCGAAGCCGCCAGGAAATTCATAATTTATGATGTCGCACTGG 1648  
 Qy 1783 AGCCTACTTAAGAGTCTTACAGCAAAAGTCTACAGCAGACCCAGATAGTTGCTGTCT 1842  
 Db 1649 AACTTATGTGAGCAATGGATGATGTTGTGGCTCAGATCCCAAACTTATATTATGCT 1708  
 Qy 1843 GTTGTCAAGTAATCGGAAGCAATAATGATGCTATTAATAAATACCTGTGTACAGATTG 1902  
 Db 1709 CGTACCAATGATTAACGCCGGAAGATCTATCATCAATCAAAAGAGAGGATACGTTGACAG 1768  
 Qy 1903 CCCTACCCCAAGTCACTGTGTGTCGCGGACCTTAGGCAACAGCAAACTGTCATGGC 1962  
 Db 1769 GCGGTGCCAACTCA---AGTTGTGACCTTTAAACGACCAAGAACCGTAGCCTTATGAG 1825

Qy 1963 CATTGCTACAAAGATTGCGCTACAGATGAATGCAAGATGGGAGGAGAGCTCTGGAGGT 2022  
 Db 1826 CATTGCGCACCAAAATAGCAATCCAACTGAATTTGCAAGTTGGGATATACACCTGGATGAT 1885  
 Qy 2023 GGACATCCCCCTCAAGCTCGTGTATGATCGTTGGCATCGATTTGTTACCATGATGACAGC 2082  
 Db 1886 CGAACTACCTTGTCTCGGACTGATGACAAATTTGGCTTTGACATTTGCAAGAGCACACGAGA 1945  
 Qy 2083 TGGCGGAGGTCAATCGCAGGATTTTGGCCAGCATCAATGAAGGAGTACCCCGCTGTT 2142  
 Db 1946 TCGGAAGAGGGCTACGAGCATTTGATTGCTCAATGATCTACAGCAAACTCCAGCTA 2005  
 Qy 2143 CTCACCTCGATATTTCAAGGATAGAGACAGAGCTGGTAGATGGGTCTAAAGTCTCGCT 2202  
 Db 2006 CTTACAGCAGTCAAGGAGTGGAGCGCTTTGATGTTGCTGCTAACACCTTTTGGCCGAT 2065  
 Qy 2203 GCAAGC---GGCTCTGAGGGCTTTGGAATAGCTGCAATGAGGTACATGCCAGCGGATCAT 2259  
 Db 2066 GATAGCAAGGCCCTGCGCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGT 2125  
 Qy 2260 CGTGTACCGCATGCGGTAGGAGACGGCCAGCTGAAACACTGGTGAACCTACGAAGTGC 2319  
 Db 2126 ATTTTATCGAGCGGTGAGCTCCGGCTCTCTAAAGCAGCTTTTGAATTTGAAGTCAA 2185  
 Qy 2320 ACAGTTTTTGGATTTCTTAAATCCATTTGGTAGAGTTTACAACTAGANTAAACGGTAAT 2379  
 Db 2186 GGACATCATTTGAGAAGTTGAAACTGAACTAGCCCGCTCCAGCTAAGCCCAACCGCAAT 2245  
 Qy 2380 TGTGTTGAAGAAAGAGTGAACACCAAGATTTTGTCTCAGTCTGGAGGAAGACTTCCAGAA 2439  
 Db 2246 AGCT---TATATTGTTGTTAAACAGATCCATGAACACCGCTTCTTCTCAACGACAAA 2302  
 Qy 2440 TCCACTTCTCGGAAGCTTATGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2499  
 Db 2303 TCTTCCGCTGTTACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2362  
 Qy 2500 TATGTTGAGCGAGCTGTGAGAAAGTGTGATGTTTCTCCACACATTTACAAATGTCATCTA 2559  
 Db 2363 TCTGTTCTCGCAACAAGTTCTGTCAGGTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTT 2422  
 Qy 2560 TGACAAACAGCGGCTGAGCCAGACACATACAGCGCTTGAACCTCAAGCTGTGCCACAT 2619  
 Db 2423 TAGCAGATCGGTTCTTCCCGGAGAAATGCAAAACTTACGTACAAAGTGTGCCACTT 2482  
 Qy 2620 CTATTCAACTGCCAGGTGTCTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2679  
 Db 2483 GTACTACAAATTTGTTGCGGACACACAGCTGCGAGGTTTGGCCAGTACGCTAAGAAGCT 2542  
 Qy 2680 GGTCTTTTCTGTTGGCCAGAGTATTCACAGAGGCAAAATCTGTCTGTCAAAACCGCT 2739  
 Db 2543 AGTACCTCTGTTGGTACAACTTGTGACTTATTCGCAAAACCGCTCGAAAGAAAGTT 2602  
 Qy 2740 TTACTACCTCTAA 2752  
 Db 2603 TTATTATCTATAA 2615

RESULT 11  
 ; Sequence 2, Application US/09774434  
 ; Patent No. US20020137044A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tang, Y. Tom  
 ; APPLICANT: Zhou, Ping  
 ; APPLICANT: Goodrich, Ryle  
 ; APPLICANT: Liu, Chenghua  
 ; APPLICANT: Asundi, Vinod  
 ; APPLICANT: Wehrman, Tom  
 ; APPLICANT: Yang, Yonghong  
 ; APPLICANT: Drmanac, Radoje T.  
 ; TITLE OF INVENTION: NO. US20020137044A1el Nucleic Acids and  
 ; TITLE OF INVENTION: Polypeptides

```

FILE REFERENCE: 787CIP2G
CURRENT APPLICATION NUMBER: US/09/774,434
CURRENT FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 09/496,914
PRIOR FILING DATE: 2000-02-03
NUMBER OF SEQ ID NOS: 6
SOFTWARE: pf_genes Version 2.0
SEQ ID NO 2
LENGTH: 2914
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(2583)
US-09-774-434-2

```

```

Query Match 3.2%; Score 111.8; DB 10; Length 2914;
Best Local Similarity 50.5%; Pred. No. 1.3e-15;
Matches 306; Conservative 0; Mismatches 288; Indels 12; Gaps 1;

```

```

QY 2093 TCAATCGAGATTTGTTCCAGATCAATGAGGATGACCGCGTGTCTCAGCGCTGC 2152
DB 1831 TCTATTGCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1890
QY 2153 ATATTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2212
DB 1891 AGAGTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1950
QY 2213 CTGAGGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2272
DB 1951 CTATTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2010
QY 2273 GGGCGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2332
DB 2011 GATTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2070
QY 2333 TGTCTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2392
DB 2071 GCGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2130
QY 2393 AGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2440
DB 2131 AGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2190
QY 2441 CCACTTCTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2500
DB 2191 ATCCAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2250
QY 2501 ATGCTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2560
DB 2251 CTCTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2310
QY 2561 GACAACTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2620
DB 2311 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2370
QY 2621 TATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2680
DB 2371 TACCTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2430
QY 2681 GCTTTT 2686
DB 2431 GCATTT 2436

```

```

RESULT 12
US-09-815-343-3
Sequence 3, Application US/09815343
Patent No. US20010055596A1
GENERAL INFORMATION:
APPLICANT: Meagher, Madeleine

```

```

APPLICANT: Xu, Jiangchun
APPLICANT: King, Gordon E.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.504
CURRENT APPLICATION NUMBER: US/09/815,343
CURRENT FILING DATE: 2001-03-22
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 444
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(444)
OTHER INFORMATION: n = A,T,C or G
US-09-815-343-3

```

```

Query Match 2.8%; Score 95.6; DB 10; Length 444;
Best Local Similarity 66.2%; Pred. No. 2.5e-12;
Matches 137; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

```

```

QY 1824 CCAGATAGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 1883
DB 122 CCAGATAGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 181
QY 1884 AATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1943
DB 182 AATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 241
QY 1944 AACAGCAATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2003
DB 242 AACAGCAATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 301
QY 2004 GAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2030
DB 302 GAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 328

```

```

RESULT 13
US-09-864-761-32092
Sequence 32092, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665

```

;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 32092  
;; LENGTH: 213  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AP000359.1  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.1  
;; OTHER INFORMATION: EST HUMAN HIT: BF246708.1, EVALUE 5.00e-05  
;; OTHER INFORMATION: NT HIT: g10946611, EVALUE 8.00e-06  
;; OTHER INFORMATION: SWISSPROT HIT: O04379, EVALUE 2.70e-02  
US-09-864-761-32092

Query Match 2.3%; Score 78.6; DB 10; Length 213;  
Best Local Similarity 67.2%; Pred. No. 1.2e-08;  
Matches 127; Conservative 0; Mismatches 59; Indels 3; Gaps 1;  
QY 1832 GTTGCTCTCTCTCAAGTAATCGGAAGACAAATACGATGCTATTAAAAATACCTG 1891  
Db 28 GTGATTGTATCTCTCCCAATGATGACAAAGTAGATATGACAGCATAAAGATACCTA 87  
QY 1892 TGTACAGATTGCCCTPACCCCAAGTCAGTGTGTGGTGGCCGCAACCTTAGGCAACACGAA 1951  
Db 88 TGTACCAATGCCCAATTCGAAGCAGGTGTGTGTGGAAGACCTTAGAAGAAAGTCCA- 146  
QY 1952 ACTGTCATGGCCANTGCTACAAAGATTGCCCTACAGATGAATGCAAGATGGGAGGAG 2011  
Db 147 --GGCAAGACATCGTCACCAAGATTGCCCAAGATGAATTGCAAGATGGGAGGAGCC 204  
QY 2012 CTCTGGAGG 2020  
Db 205 CTCTGGAAG 213

RESULT 14  
US-09-864-761-4964  
; Sequence 4964, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aeonica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6

;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 4964  
;; LENGTH: 364  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AP000359.1  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.98  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.92  
US-09-864-761-4964

Query Match 2.0%; Score 68; DB 10; Length 364;  
Best Local Similarity 62.2%; Pred. No. 3.9e-06;  
Matches 107; Conservative 0; Mismatches 65; Indels 0; Gaps 0;  
QY 889 TCCAAGTCACAGCTTCGTTGATTTGGCTTCACTACTTCCATCTTCCATCTTCAAGTATGGA 948  
Db 193 TTCTAGTACCAGTTTGGAAATCTGGCTTGGTTATGTTACTTCTTCTTCAATACGAAA 252  
QY 949 CAGCATCATGCTCTGCATGACGTTAGCCATAAAGTCTTCCAAAGTGAAGACTGTTTGGGA 1008  
Db 253 CAGCATTACCCTCTGTGCGATGTGAGCCACAACTGCTCCGATAGAACTGCTTATGA 312  
QY 1009 TTTCTATGTTCAACTTTTATCATCATCAGACAGAGAAATTTCAAGAACAA 1060  
Db 313 TTTCTAAGAGAACATCTGCCAGGCCACAGAGGAAACATCCGAGAGGAA 364

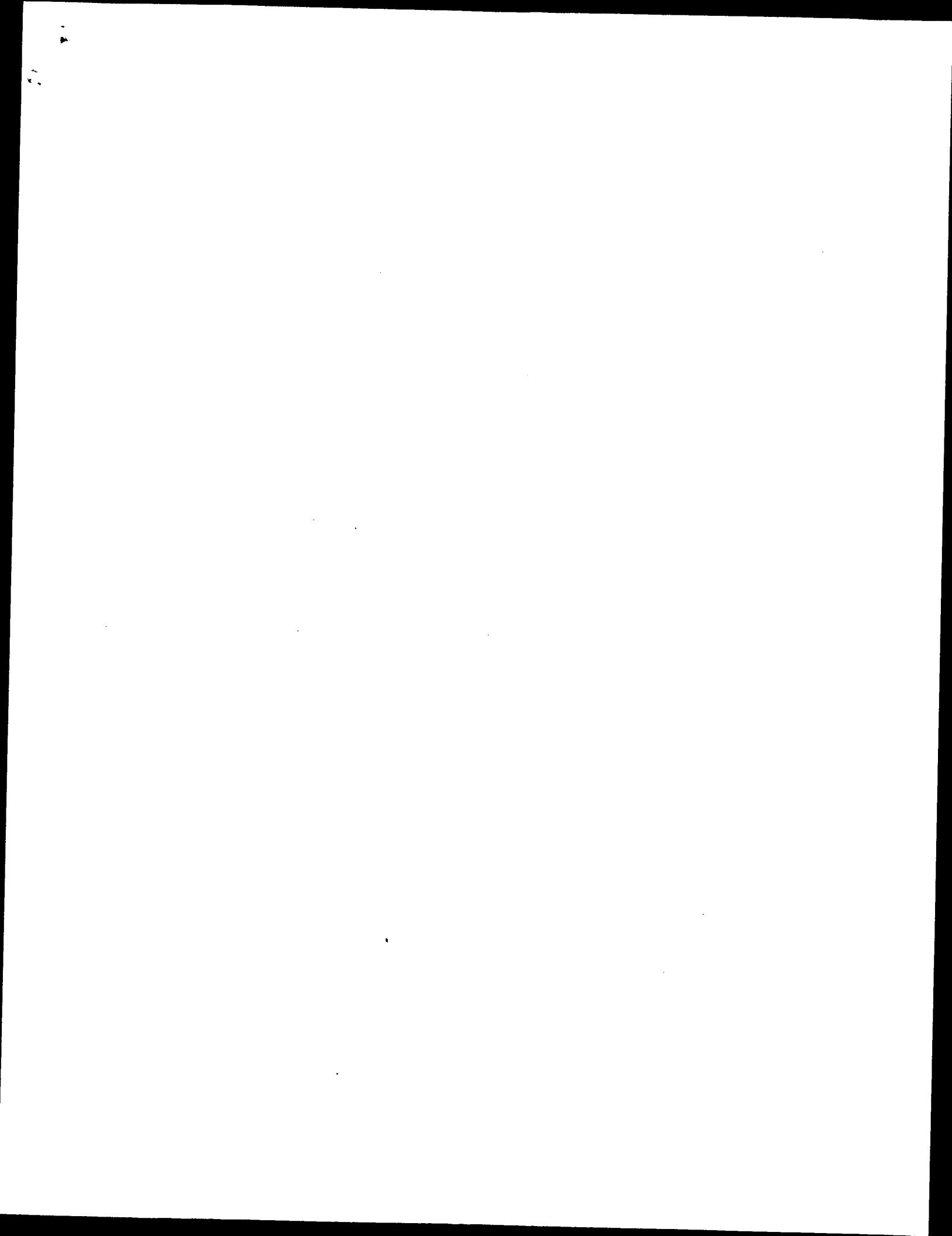
RESULT 15  
US-09-960-352-3400  
; Sequence 3400, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960,352  
; CURRENT FILING DATE: 2001-09-24

NUMBER OF SEQ ID NOS: 15112  
 SEQ ID NO 3400  
 LENGTH: 446  
 TYPE: DNA  
 ORGANISM: Bos taurus  
 OTHER INFORMATION: Clone ID: 15-LIB3058-052-Q1-K1-D11  
 US-09-960-352-3400

Query Match 1.9%; Score 67; DB 10; Length 446;  
 Best Local Similarity 48.3%; Pred. No. 7.2e-06;  
 Matches 187; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

QY 2768 GCAGCGCTTTCTTTGAAATGACTTGGATTTTAACTTTTAACTTTT 2827  
 DB 14 GCGGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 73  
 QY 2828 TTTAACTGTTATCTTCTGAGAACTTGGAGGGATAGAGATCTAGCATTTAT 2887  
 DB 74 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 133  
 QY 2888 TTCTAGATGCTATTCACGGCTTCTTATTTATATGTAATAATAGATTTTAT 2947  
 DB 134 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 193  
 QY 2948 TTATCTTCTGTTTCATAGATATTTTGTGAGCATTTTGTGTTTGAAGAAAG 3007  
 DB 194 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 253  
 QY 3008 TGAATAGATCTGTAAGATTAACAAGACTCTGAGAGTATTGAAATGTTTGA 3067  
 DB 254 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 313  
 QY 3068 GATTACTTAAAGTACTTTCAGAGTGAAGCAAGTCTTATTAACCTATTAATT 3127  
 DB 314 TATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 373  
 QY 3128 TATTTTGAATACCTGTTTGAATT 3154  
 DB 374 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 400

Search completed: January 9, 2003, 23:04:49  
 Job time : 137 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 18:54:20 ; Search time 3076 Seconds  
(without alignments)  
18280.471 Million cell updates/sec

Title: US-09-873-737a-5  
Perfect score: 3472  
Sequence: 1 gggcgcaattggccgcagcgt.....taataataattgctccccc 3472

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_esthum:\*  
4: em\_esthum:\*  
5: em\_estcov:\*  
6: em\_estcpl:\*  
7: em\_estcpl:\*  
8: em\_estcpl:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_est3:\*  
12: gb\_est4:\*  
13: gb\_est5:\*  
14: em\_estfun:\*  
15: em\_estfun:\*  
16: em\_estfun:\*  
17: gb\_est6:\*  
18: em\_gsa\_hum:\*  
19: em\_gsa\_hum:\*  
20: em\_gsa\_hum:\*  
21: em\_gsa\_hum:\*  
22: em\_gsa\_hum:\*  
23: em\_gsa\_hum:\*  
24: em\_gsa\_hum:\*  
25: em\_gsa\_hum:\*  
26: em\_gsa\_hum:\*  
27: em\_gsa\_hum:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	736.8	21.2	818	BG773137 602721862
2	730.2	21.0	779	B1560710 603205474
3	689.2	19.9	1092	B1463482 603200456
4	672.2	19.4	882	B1459464 603200456
5	637.2	18.4	715	BG721550 602695155
6	633	18.2	666	B1463897 603202635

7	626.2	18.0	731	B1458664
8	621.4	17.9	640	AL705301
9	621	17.9	622	AL705072
10	580.6	16.7	718	BQ35882 AGNCOURT
11	574.6	16.5	874	BF66315 602119233
12	560.6	16.1	581	BF66618 602696127
13	523.2	15.1	539	B1463615
14	521	15.0	744	B1459142
15	515.2	14.8	577	BF080918
16	512.1	14.8	527	BF751055
17	502.4	14.5	1495	BF649861
18	492.8	14.2	605	BF704986
19	492.6	14.2	705	BF246708
20	471.2	13.6	533	BF080923
21	468.2	13.5	830	BF699368
22	462.4	13.0	463	AA639672
23	450.4	12.8	482	AM189518
24	443.4	12.7	474	AL705611
25	442.2	12.7	806	BG724366
26	440.4	12.7	462	AA669938
27	434.4	12.5	454	AL125224
28	415.8	12.0	454	AL044133
29	414.2	11.9	565	BG101107
30	403.4	11.6	413	AL041473
31	387.8	11.2	420	BQ374910
32	387.8	11.1	477	BQ374910
33	387	11.1	587	AA430311
34	373.4	10.8	935	BF247005
35	373.4	10.8	935	BF247005
36	345.2	9.9	485	BG178140
37	339	9.8	971	BB144499
38	328.2	9.5	349	AA904973
39	326.8	9.4	872	B1560401
40	287.2	8.3	609	BQ481304
41	274.4	7.9	559	BW728018
42	274.4	7.9	300	A1538315
43	265.4	7.6	286	A1698821
44	263.8	7.5	512	BM680592
45	259.2	7.5	617	BQ40023

ALIGNMENTS

RESULT 1  
BG773137 818 bp mRNA EST 15-MAY-2001  
LOCUS 602721862P1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:4838918 5',  
DEFINITION  
mRNA sequence.  
ACCESSION BG773137  
VERSION BG773137.1 GI:14083790  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 818)  
NIH-MGC http://mgs.nci.nih.gov/  
NATIONAL INSTITUTES OF HEALTH, Mammalian Gene Collection (MGC)  
UNPUBLISHED (1999)  
CONTACT: Robert Strauberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
Tohshiyuki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLNL0773 row: k column: 15  
High quality sequence stop: 789.  
Location/Qualifiers

```

source
1. 818
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4838918"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/notes="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 245 a 174 c 198 g 201 t
ORIGIN

Query Match 21.2%; Score 736.8; DB 12; Length 818;
Best Local Similarity 98.2%; Pred. No. 1.6e-146;
Matches 798; Conservative 0; Mismatches 8; Indels 7; Gaps 5;
QY 133 AGGCGGAGGCGCGGTCCCAAGAAATAGAAAACAATGACTGGAGAGCCCGAGCCAGAGC 192
DB 6 AGGCTGAGGCGCGGTCCCAAGAAATAGAAAACAATGACTGGAGAGCCCGAGCCAGAGC 65
QY 193 CAGAGGAAGGCGCGGTCCAGAGACAGCGAGCTGGTGGGCTCCACTGCGCATCAGCA 252
DB 66 CAGAGGAAGGCGCGGTCCAGAGACAGCGAGCTGGTGGGCTCCACTGCGCATCAGCA 125
QY 253 ACCTGGTTATATTCAGCTTAGCGCTCAGCGCCACAGCAGAGGGGGAATTTTGGCCG 312
DB 126 ACCTGGTTATATTCAGCTTAGCGCTCAGCGCCACAGCAGAGGGGGAATTTTGGCCG 185
QY 313 TGGACGCGCAGAGGAAACAGCAGGAGGAAACAGCAAGTCTCAGAGTCCAGATCTCTGC 372
DB 186 TGGACGCGCAGAGGAAACAGCAGGAGGAAACAGCAAGTCTCAGAGTCCAGATCTCTGC 245
QY 373 TGGATTTTCAAGGATTTATCGTACGCGAGAGAGGAGTCTGCTAGAGATTTTCATGATCT 432
DB 246 TGGATTTTCAAGGATTTATCGTACGCGAGAGAGGAGTCTGCTAGAGATTTTCATGATCT 305
QY 433 TGGTGTGAATACAGGCAAGACCTAGACCATGTTTAAAGATCAAAACAGGTTCTTCAGG 492
DB 306 TGGTGTGAATACAGGCAAGACCTAGACCATGTTTAAAGATCAAAACAGGTTCTTCAGG 365
QY 493 CATTATAGTAGGTAAAGCACTAACATTTCCGGCTGACATCCCGTCCCGAGTGGGCTT 552
DB 366 CATTATAGTAGGTAAAGCACTAACATTTCCGGCTGACATCCCGTCCCGAGTGGGCTT 425
QY 553 ATATCAGTATCACAATGACTATTAACCCACTGTAGGAGCCAGAACTCCGTTCACTCT 612
DB 426 ATATCAGTATCACAATGACTATTAACCCACTGTAGGAGCCAGAACTCCGTTCACTCT 485
QY 613 TCTTTTTCACACAGAGATCTAATTTGGAAGTGTCTATGCTTTTGTAGTGAACGATATTATT 672
DB 486 TCTTTTTCACACAGAGATCTAATTTGGAAGTGTCTATGCTTTTGTAGTGAACGATATTATT 545
QY 673 TTTTACCTTAAAGACTACAGCAAGGTTACTGAGTCTTTTGTAGTGAACGATATTATT 732
DB 546 TTTTACCTTAAAGACTACAGCAAGGTTACTGAGTCTTTTGTAGTGAACGATATTATT 605
QY 733 GGATGTGAGGATACGATCACTTTTAAACAA - TGAACCTTCCACCTTACATCACCACCTTGT 791
DB 606 GGATGTGAGGATACGATCACTTTTAAACAAATGAACTTCCACCTTACATCACCACCTTGT 665
QY 792 TGCAGTTCTTATAATATTATTTTCAGGAGGC - TTTTGAATATCATGAATTTGC --- AACAA 847
DB 666 TGCAGTTCTTATAATATTATTTTCAGGAGGC TTTTGAATATCATGAATTTGC CAAACAAA 725
QY 848 ATTGGACGAAATTTATTAAACCCAAATGACCCAA - TTGATATTCCAAAGTCACAGGTTGCT 906
DB 726 TTGGGACGAAATTTATTAAACCCAAATGACCCAAATTTGATTTTCCAAAGTCACAGGTTGCT 785

```

```

QY 907 GATTGG-CCTGGCTTCACTACTTCCATCCTTC 938
DB 786 GATTGGCTTGGCTTCACTACTTCCATCCTTC 818

RESULT 2
BI560710
LOCUS
DEFINITION
603254744F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5296954 5',
mRNA sequence.
BI560710
ACCESSION
BI560710.1 GI:15448024
VERSION
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
1 (bases 1 to 779)
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11751 Row: h column: 11
High quality sequence stop: 732.
FEATURES
Location/Qualifiers
1..779
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5296954"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/notes="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 234 a 159 c 205 g 181 t
ORIGIN

```

```

Query Match 21.0%; Score 730.2; DB 13; Length 779;
Best Local Similarity 99.2%; Pred. No. 4e-145;
Matches 754; Conservative 0; Mismatches 4; Indels 2; Gaps 2;
QY 108 GGTCTGAGGTGCAAGGACCAAGGACTAGGCGAGGCGCGGTCCCAAGAAATAGAAAACAA 167
DB 20 GGTCTGAGGTGCAAGGACCAAGGACTAGGCGAGGCGCGGTCCCAAGAAATAGAAAACAA 79
QY 168 TGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAGGCGCGGTCCAGAGACAGCCGAGC 227
DB 80 TGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAGGCGCGGTCCAGAGACAGCCGAGC 139
QY 228 TGGTGGGCTTCACTGCGCATGCAACCTGGTTATTTTCAGCTTAGGCTTCAGCCCTCAC 287
DB 140 TGGTGGGCTTCACTGCGCATGCAACCTGGTTATTTTCAGCTTAGGCTTCAGCCCTCAC 199
QY 288 CAGCAGAGGGGGAATTTATTTGGCCGTGAGCGGAGAGGAAACAGCAGGAGGAAACAGCCA 347
DB 200 CAGCAGAGGGGGAATTTATTTGGCCGTGAGCGGAGAGGAAACAGCAGGAGGAAACAGCCA 259

```

348 AGTCAAGAGACTCGATATCTGCTGATTTGAGAGATTGCTGATGAGAGAGAG 407  
 260 ACTCAAGAGACTCGATATCTGCTGATTTGAGAGATTGCTGATGAGAGAGAG 319  
 408 GTCTGCTGATGAGATTTGATGATTTGCTGCTGATTTGAGAGAGAGAGAG 467  
 320 GTCTGCTGATGAGATTTGATGATTTGCTGCTGATTTGAGAGAGAGAGAG 379  
 468 AAGATCAAAAACAGGTTCTTCAAGCATTTATAGTAAGGTTAAGCACTTAACCAATTTCCGCG 527  
 380 AAGATCAAAAACAGGTTCTTCAAGCATTTATAGTAAGGTTAAGCACTTAACCAATTTCCGCG 439  
 528 TGACATCCCGTCCGAGGCGCTTATGATGATGATGATGATGATGATGATG 587  
 440 TGACATCCCGTCCGAGGCGCTTATGATGATGATGATGATGATGATGATG 499  
 588 AAGCAGAGAGAGCTCGTCTGATGATGATGATGATGATGATGATGATG 647  
 500 AAGCAGAGAGAGCTCGTCTGATGATGATGATGATGATGATGATGATG 559  
 648 ATGCTTTGATGAGATTTATTTATTTTACCTTAAGAGAGAGAGAGAGAG 707  
 560 ATGCTTTGATGAGATTTATTTATTTTACCTTAAGAGAGAGAGAGAGAG 619  
 708 TTTTAAAG 767  
 620 TTTTAAAG 679  
 768 TTCACTTCAATCAACCACTGTTGAGATTTATTTATTTTATTTTATTTT 827  
 680 TTCACTTCAATCAACCACTGTTGAGATTTATTTATTTTATTTTATTTT 738  
 828 AATATCA-TGAATTTGCAACAATTTGAGAGAGAGAGAGAGAGAGAGAG 866  
 739 AATATTTGCAACAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 778

RESULT 3  
 BI463482  
 LOCUS  
 DEFINITION 603203560F1 NIH\_MGC\_97 Homo sapiens CDNA clone IMAGE:5269311 5',  
 mRNA sequence.  
 BI463482  
 ACCESSION  
 VERSION BI463482.1 GI:15254125  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgsapb-r@mail.nih.gov  
 Tissue Procurement: Miklos Palkeovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki  
 Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed By: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Inocyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLN at:  
 http://image.llnl.gov  
 Plate: LLM11679 row: h column: 16  
 High quality sequence stop: 668.  
 Location/Qualifiers  
 1..1092  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5269311"  
 /clone\_lib="NIH\_MGC\_97"  
 /lab\_host="DH10B"  
 /note="Organ: testis; Vector: pBluescriptR (modified)

BASE COUNT 356 a 201 c 270 g 264 t 1 others  
 ORIGIN  
 Bluescript KS+; Site 1: BamHI; Site 2: SalI-XhoI (gcgcag  
 ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
 size-selected for average insert size 2.2 kb and  
 normalized to 10^5. This is a primary library enriched  
 for full-length clones and constructed using the  
 Cap-trapper method (Carninci, in preparation). Library  
 constructed by M. Brownstein (NHGRI/NHGRI, National  
 Institutes of Health). Note: this is a NIH-MGC Library."

Query Match 19.9%; Score 689.2; DB 13; Length 1092;  
 Best Local Similarity 83.5%; Pred. No. 2,1e-136;  
 Matches 903; Conservative 0; Mismatches 156; Indels 22; Gaps 10;

106 CAGGTGAGGTCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 165  
 16 CAGGAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 75  
 166 AATGACTGGAG 225  
 76 AAGACTGGAG 135  
 226 G-CTGAGTGGCTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCC 284  
 136 GACTGAGTGGCTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCC 195  
 285 CACGAG 344  
 196 CACGAG 255  
 345 CCAAGTCAAG 404  
 256 CCAAGTCAAG 315  
 405 GAGTGGTGAAG 464  
 316 GAGTGGTGAAG 375  
 465 TTAAGAGATCAAAAACAGGTTCTTCAAGCATTTATAGTAAGTGAAGTGAAG 524  
 376 TTAAGAGATCAAAAACAGGTTCTTCAAGCATTTATAGTAAGTGAAGTGAAG 435  
 525 GAGTGAATCCCGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 584  
 436 GAGTGAATCCCGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 495  
 585 TGAAG 644  
 496 TGAAG 555  
 645 GTGATGCTTTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 704  
 556 GTGATGCTTTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 615  
 705 AAGTTTAAAG 764  
 616 AAGTTTAAAG 675  
 765 AAGTTTAAAG 824  
 676 AAGTTTAAAG 732  
 825 TGAAGATCATGATTTGCAACAATTTGAGAGAGAGAGAGAGAGAGAGAGAG 881  
 733 TGAAGATCATGATTTGCAACAATTTGAGAGAGAGAGAGAGAGAGAGAGAG 792  
 882 TGAAGATCATGATTTGCAACAATTTGAGAGAGAGAGAGAGAGAGAGAGAG 933  
 793 TGAAGATCATGATTTGCAACAATTTGAGAGAGAGAGAGAGAGAGAGAGAG 852  
 934 CTTGAGTGAAG 989





Db 125 AGGACAGCGCAGCTGGTGGCTCCACTGCCAGTCAGCAACCTGGTTATATTACGCTA 184  
 Qy 273 GGCCTCAGCCGCCACACAGAGGGGGAATTTATTTGGCCGTGGAGCGGAGAGGAACAG 332  
 Db 185 GGCCTCAGCCGCCACACAGAGGGGGAATTTATTTGGCCGTGGAGCGGAGAGGAACAG 244  
 Qy 333 CAGGAGGAACAGCCAAAGTCACAGAGCTCCAGATATCTGCTGGATTTTCAGAGTTATCGN 392  
 Db 245 CAGGAGGAACAGCCAAAGTCACAGAGCTCCAGATATCTGCTGGATTTTCAGAGTTATCGT 304  
 Qy 393 TAGCAGAGAGAGGAGGCTCGTCTAGAGATTTTCATGATCTTGTGTGAATACAGGCGAGA 452  
 Db 305 TAGCAGAGAGAGGAGGCTCGTCTAGAGATTTTCATGATCTTGTGTGAATACAGGCGAGA 364  
 Qy 453 ACCTAGAGCAGGTTTAAAGATCAAAACAGAGCTTCTTCAGGCATATATAGTTAAGCA 512  
 Db 365 ACCTAGAGCAGGTTTAAAGATCAAAACAGAGCTTCTTCAGGCATATATAGTTAAGCA 424  
 Qy 513 CTAACCACTTTCCGGCTGACATCCCGTCCCGAGTGGGCTTTATATCATGATCATCATTTGACT 572  
 Db 425 CTAACCACTTTCCGGCTGACATCCCGTCCCGAGTGGGCTTTATATCATGATCATCATTTGACT 484  
 Qy 573 ATAAACCACTGATGAA- GCCAGAGACTCCGTTTTCAGCTCTTCTTTTCAACAGAGAT 631  
 Db 485 ATAAACCACTGATGAAACGCCAGAGACTCCGTTTTCAGCTCTTCTTTTCAACAGAGAT 544  
 Qy 632 CTAATTCGAAAGTGTATGCTTTTGTAGGAAACGATATATTTTACCTAAAGACTACAG 691  
 Db 545 CTAATTCGAAAGTGTATGCTTTTGTAGGAAACGATATATTTTACCTAAAGACTACAG 604  
 Qy 692 CAAAGGTTACTGAAGTTTGTAGTAAACCCGGAATGGAGAGATGTAGGATAACGATC 751  
 Db 605 CAAAGGTTACTGAAGTTTGTAGTAAACCCGGAATGGAGAGATGTAGGATAACGATC 664  
 Qy 752 AC 753  
 Db 665 AC 666

RESULT 7  
 BI458664  
 LOCUS 603199792F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:5265838 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BI458664  
 VERSION BI458664.1 GI:15249320  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 731)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs@mail.nih.gov  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
 Toshinuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLNL1670 row: g column: 23  
 High quality sequence stop: 671.  
 Location/Qualifiers  
 1. .731  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5265838"  
 /clone\_lib="NIH\_MGC\_97"

FEATURES  
 source

/lab\_host="DH10B"  
 /note="Organ: testis; Vector: pBluescriptR (modified  
 pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag  
 ); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
 size-selected for average insert size 2.2 kb and  
 normalized to ROT 5. This is a primary library enriched  
 for full-length clones and constructed using the  
 Cap-trapper method (Carninci, in preparation). Library  
 constructed by M. Brownstein (NHGRI/NHGRI, National  
 Institutes of Health). Note: this is a NIH\_MGC Library."  
 BASE COUNT 214 a 157 c 201 g 159 t  
 ORIGIN

Query Match 18.0%; Score 626.2; DB 13; Length 731;  
 Best Local Similarity 97.5%; Pred. No. 5.8e-123;  
 Matches 667; Conservative 0; Mismatches 14; Indels 3; Gaps 3;  
 Qy 93 GGGGCGCCCGCGGAGGCTCGAGGTGCAAGCAGCAGGACTAGGGCGGAGGCGGTCCA 152  
 Db 18 GGGGCTGTTGGCTCGGGCTGAGGTGCAAGCAGCAGGACTAGGGCGGAGGCGGTCCA 77  
 Qy 153 AGAAATAGAAAAACAATCACTGGGAGAGCCCGAGCAGAGGAGGAGGCGCGGTCC 212  
 Db 78 AGAAATAGAAAAACAATCACTGGGAGAGCCCGAGCAGAGGAGGAGGCGCGGTCC 137  
 Qy 213 AGGAGACAGCGAGCTGGTGGCTCCACTGCGAGTCAGCAACCTGGTTATATTCAGCCTA 272  
 Db 138 AGGAGACAGCGAGCTGGTGGCTCCACTGCGAGTCAGCAACCTGGTTATATTCAGCCTA 197  
 Qy 273 GGCCTCAGCCGCCACACAGCAGAGGGGGAATTTATTTGGCCGTGGAGCGGAGAGGAACAG 332  
 Db 198 GGCCTCAGCCGCCACACAGCAGAGGGGGAATTTATTTGGCCGTGGAGCGGAGAGGAACAG 257  
 Qy 333 CAGGAGGAACAGCCAAAGTCACAGAGCTCCAGATATCTGCTGGATTTTCAGAGTTATCGN 392  
 Db 258 CAGGAGGAACAGCCAAAGTCACAGAGCTCCAGATATCTGCTGGATTTTCAGAGTTATCGT 317  
 Qy 393 TAGCAGAGAGAGGAGGCTCGTCTAGAGATTTTCATGATCTTGTGTGAATACAGGCGAGA 452  
 Db 318 TAGCAGAGAGAGGAGGCTCGTCTAGAGATTTTCATGATCTTGTGTGAATACAGGCGAGA 377  
 Qy 453 ACCTAGAGCAGGTTTAA-AGAATCAAAACAGGTTCTTCAGGCATATATAGTTAAGTTAAGC 511  
 Db 378 ACCTAGAGCAGGTTTAAAGATCAAAACAGGTTCTTCAGGCATATATAGTTAAGTTAAGC 437  
 Qy 512 ACTAACCATTTCCGGCTGACATCCCGTCCCGAGTGGGCTTTATATCATGATCATCATTTGAC 571  
 Db 438 ACTAACCATTTCCGGCTGACATCCCGTCCCGAGTGGGCTTTATATCATGATCATCATTTGAC 497  
 Qy 572 TATAACCCACTGATGGAAGCGCAGAGACTCCGTTACGCTCTCTTTTCAACAGAGAT 631  
 Db 498 TATAACCCACTGATGGAAGCGCAGAGACTCCGTTACGCTCTCTTTTCAACAGAGAT 557  
 Qy 632 CTAATTCGAAAGTGTATGCTTTTGTAGTGAACGATATATTTTACCTAAAGACTACAG- 690  
 Db 558 CTAATTCGAAAGTGTATGCTTTTGTAGTGAACGATATATTTTACCTAAAGACTACAG 617  
 Qy 691 GCAAAAGGTTACT- GAAAGTTTGTAGTGAACCGGAGTGGAGAGATGTAGAGATAACGA 749  
 Db 618 GCAAAAGGTTACTGGAAGTTTGTAGTGAACCGGAGTGGAGAGATGTAGAGATAACGA 677  
 Qy 750 TCACCTTTAAACAAATGACATTCAC 773  
 Db 678 TCACCTTTAAACAAATGACATTCAC 701

RESULT 8  
 AL705301  
 LOCUS  
 DEFINITION DKFP686G1235\_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone  
 ACCESSION DKFP686G1235 5', mRNA sequence.  
 VERSION AL705301  
 AL705301.1 GI:19688656

AL705301 640 bp mRNA linear EST 22-MAR-2002  
 DKFP686G1235\_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone  
 DKFP686G1235 5', mRNA sequence.  
 AL705301  
 AL705301.1 GI:19688656

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS 1 (bases 1 to 640)  
Ostenwaelder, B., Obermaier, B., Mewes, W., Mewes, H. W., Weill, B. and Wiemann, S.  
TITLE EST (Ostenwaelder, B., Obermaier, B., Mewes, H. W., Weill, B. and Wiemann, S.)  
JOURNAL Unpublished (2001)  
COMMENT Contact: Ostenwaelder B  
MIPS  
Am Klopferpit 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de  
sequenced by Medigenomix (Martinsried/Germany) within the CDNA sequencing consortium of the German Genome Project. No si sequence available.  
This clone (DKFZ686G1235) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
Location/Qualifiers  
1..640  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZ686G1235"  
/clone\_lib="686 (synonym: hlcc3)"  
/tissue\_type="human skeletal muscle"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Vector: pTriplEx2; Site\_1: sf11A; Site\_2: sf11B; cDNA-collection"  
CDNA-collection

BASE COUNT 190 a 139 c 174 g 137 t  
ORIGIN

Query Match 17.9%; Score 621.4; DB 9; Length 640;  
Best Local Similarity 98.9%; Pred. No. 6.2e-122;  
Matches 625; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 108 GGTCTGAGGTGCAAGACCAAGACTAGGGCGAGGCGCGCTCCAGAAATGAAAAACA 167  
DB 9 GGGCTAGGGTCAAGACCAAGACTAGGGCGAGGCGCGCTCCAGAAATGAAAAACA 68

QY 168 TGACTGGAGAGCCCGACCAAGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 227  
DB 69 TGACTGGAGAGCCCGACCAAGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 128

QY 228 TGGTGGGCTCCATGCGACAGTCAAGCACTGGTTATATTCAGCTAGGCTCAGCGCCAC 287  
DB 129 TGGTGGGCTCCATGCGACAGTCAAGCACTGGTTATATTCAGCTAGGCTCAGCGCCAC 188

QY 288 CAGCAGAGGGGGAATTTATTTGGCCGTGAGCGGAGAGGAGGAGGAGGAGGAGGAGG 347  
DB 189 CAGCAGAGGGGGAATTTATTTGGCCGTGAGCGGAGGAGGAGGAGGAGGAGGAGGAGG 248

QY 348 AGTCAACAGGACTCCAGATATCTGTGATTTCAAGAGTTATCGTATGAGAGAGAGAG 407  
DB 249 AGTCAACAGGACTCCAGATATCTGTGATTTCAAGAGTTATCGTATGAGAGAGAGAG 308

QY 408 GTGCTGTAGAGATTTTCATGATCTTGTGTGATTAACAAGGCAAGACCTAGACATGTTA 467  
DB 309 GTGCTGTAGAGATTTTCATGATCTTGTGTGATTAACAAGGCAAGACCTAGACATGTTA 368

QY 468 AAGAAATAAAAAGAGTTCTTCAAGGATTTATAGAGGTTAAGCACTAATTCATTCGCG 527  
DB 369 AAGAAATAAAAAGAGTTCTTCAAGGATTTATAGAGGTTAAGCACTAATTCATTCGCG 428

QY 528 TGACATCCCGTCCAGAGTGGGCTTATATCAGTATCATATTGACTAATCCCACTGATGG 587  
DB 429 TGACATCCCGTCCAGAGTGGGCTTATATCAGTATCATATTGACTAATCCCACTGATGG 488

QY 588 AGCCAGAAAGACTCCGTTAGCTCTTTTTCACAGCAAGATTAATTTGAAAGTGTTC 647  
DB 489 AAGCCAGAAAGACTCCGTTAGCTCTTTTTCACAGCAAGATTAATTTGAAAGTGTTC 548

QY 648 ATGCTTTTGAATGAGAGATTTATTTTACCTTAAAGACTTACAGCAAGGTTACGAG 707  
DB 549 ATGCTTTTGAATGAGAGATTTATTTTACCTTAAAGACTTACAGCAAGGTTACGAG 608

QY 708 TTTTGTAGTAAACCCGGAATGAGAGATGTG 739  
DB 609 TTTTGTAGTAAACCCGGAATGAGAGATGTG 640

RESULT 9  
AL705072 622 bp mRNA linear EST 22-MAR-2002  
LOCUS  
DEFINITION DKFZ686E0634.F1 686 (synonym: hlcc3) Homo sapiens cDNA clone  
ACCESSION AL705072  
VERSION AL705072.1 GI:19688427  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS 1 (bases 1 to 622)  
Ostenwaelder, B., Obermaier, B., Mewes, W., Mewes, H. W., Weill, B. and Wiemann, S.  
TITLE EST (Ostenwaelder, B., Obermaier, B., Mewes, W., Weill, B. and Wiemann, S.)  
JOURNAL Unpublished (2001)  
COMMENT Contact: Ostenwaelder B  
MIPS  
Am Klopferpit 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de  
sequenced by Medigenomix (Martinsried/Germany) within the CDNA sequencing consortium of the German Genome Project. No si sequence available.  
This clone (DKFZ686E0634) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
Location/Qualifiers  
1..622  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZ686E0634"  
/clone\_lib="686 (synonym: hlcc3)"  
/tissue\_type="human skeletal muscle"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Vector: pTriplEx2; Site\_1: sf11A; Site\_2: sf11B; cDNA-collection"  
CDNA-collection

BASE COUNT 187 a 134 c 168 g 133 t  
ORIGIN

Query Match 17.9%; Score 621; DB 9; Length 622;  
Best Local Similarity 99.8%; Pred. No. 7.6e-122;  
Matches 621; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 119 CAAGGACCAAGACTAGAGGCGAGGCGAGCGGTCCAGAAATGAAAAACAATGACTGGAG 178  
DB 1 CAAGGACCAAGACTAGAGGCGAGGCGAGGCGGTCCAGAAATGAAAAACAATGACTGGAG 60

QY 179 GCCGAGGCAAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 238  
DB 61 GCCGAGGCAAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120

QY 239 ACTGCAGTCAAGCAACTGTTATATTCAGCTTATAGGCTTACAGCCGACCAAGAGAGG 298  
DB 121 ACTGCAGTCAAGCAACTGTTATATTCAGCTTATAGGCTTACAGCCGACCAAGAGAGG 180



299 GAATTTATTTGGCGCTGGACGCGCAGAGGACACAGCAGGAGGACAGCAAGTCAACAGGA 358  
 181 GAATTTATTTGGCGCTGGACGCGCAGAGGAGAACACAGCAGGAGGAGCAAGTCAACAGGA 240  
 359 CTCAGATATCTGCTGGATTTTCAGAGGTTATCGNTAGCAGAGAGAGGAGGTCGCTGTAGA 418  
 241 CTCAGATATCTGCTGGATTTTCAGAGGTTATCGTTAGCAGAGAGAGGAGGTCGCTGTAGA 300  
 419 GATTTTCATGATCTGCTGGATTTTCAGAGGTTATCGTTAGCAGAGAGAGGAGGTCGCTGTAGA 478  
 301 GATTTTCATGATCTGCTGGATTTTCAGAGGTTATCGTTAGCAGAGAGAGGAGGTCGCTGTAGA 360  
 479 ACAGGTTCTTCAGGCAATATAGTAGTAAAGGTTAAGCACTAACCAATTTCCGGCTGACATCCCGT 538  
 361 ACAGGTTCTTCAGGCAATATAGTAGTAAAGGTTAAGCACTAACCAATTTCCGGCTGACATCCCGT 420  
 539 CCCAGTGGGCTTTATATCAGTATCAGTATCAGTATCAGTATCAGTATCAGTATCAGTATCAGTAT 598  
 421 CCCAGTGGGCTTTATATCAGTATCAGTATCAGTATCAGTATCAGTATCAGTATCAGTATCAGTAT 480  
 599 CTCGTTTCAGCTCTCTTTTTCACACGAAAGATCTAATTCGAAAGGTCATGCTTTTGTAT 658  
 481 CTCGTTTCAGCTCTCTTTTTCACACGAAAGATCTAATTCGAAAGGTCATGCTTTTGTAT 540  
 659 GGAACGATATATTTTACCTTAAGACTACAGCAAAAGGTTACTGAAGTTTGTAGTAAG 718  
 541 GGAACGATATATTTTACCTTAAGACTACAGCAAAAGGTTACTGAAGTTTGTAGTAAG 600  
 719 ACCCGAATGGAGGATGTGA 740  
 601 ACCCGAATGGAGGATGTGA 622

RESULT 10  
 BQ435882  
 LOCUS  
 DEFINITION BQ435882 718 bp mRNA linear EST 24-MAY-2002  
 AGENCOURT\_7897283 NIH\_MGC\_82 Homo sapiens cDNA clone IMAGE:6103498  
 5', mRNA sequence.

ACCESSION BQ435882  
 VERSION BQ435882.1 GI:21174947  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 718)  
 NIH-MGC http://imgc.nhl.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs@mail.nih.gov

Tissue Procurement: CLONTECH  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLCM2339 row: n column: 11  
 High quality sequence stop: 518.  
 Location/Qualifiers

1. .718  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6103498"  
 /clone\_lib="NIH MGC 82"  
 /lab\_host="DH105 (T1 phage-resistant)"  
 /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site: 1:  
 SfiI (ggccgctcgcc); Site 2: SfiI (ggccattatgcc); 5' and  
 3' adaptors were used in cloning as follows: 5' adaptor  
 sequence: 5'-CACGCCATATATGCC-3' and 3' adaptor sequence:  
 5'-ATTCTAGAGCGCGCGCGGACATG-dt(30)BN-3' (where B = A,  
 C, or G and N = A, C, G, or T). Average insert size

1.35 kb (range 0.9-4.0 kb). 14/15 colonies contained  
 inserts by PCR. This library was enriched for full-length  
 clones and was constructed by Clontech Laboratories (Palo  
 Alto, CA)."  
 BASE COUNT 192 a 137 c 140 g 249 t  
 ORIGIN

Query Match 16.7%; Score 580.6; DB 14; Length 718;  
 Best Local Similarity 98.3%; Pred. No. 3e-113;  
 Matches 640; Conservative 0; Mismatches 4; Indels 7; Gaps 5;

QY 2504 GTGAGCCAGGCTGTGAGAGTGTAGTGTCTCCACACATTAACAATGTCATCTATGAC 2563  
 Db 5 GTGAGCCAGGCTGTGAGAGTGTAGTGTCTCCACACATTAACAATGTCATCTATGAC 64

QY 2564 AACAGCGGCTGAAGCCAGACACATACAGCGCTTGACCTACAGCTGTGCCACATCTAT 2623  
 Db 65 AACAGCGGCTGAAGCCAGACACATACAGCGCTTGACCTACAGCTGTGCCACATCTAT 124

QY 2624 TACAACTGCGCAGGTGTCAATTCGTTCTCTCTCCAGTACGCCCAAGCTGGCT 2683  
 Db 125 TACAACTGCGCAGGTGTCAATTCGTTCTCTCTCCAGTACGCCCAAGCTGGCT 184

QY 2684 TTTCTTTGTCGCGAGATTTTCACAGAGAGCCAAATCTGTCACTGTCAAAACCGCTTTTAC 2743  
 Db 185 TTTCTTTGTCGCGAGATTTTCACAGAGAGCCAAATCTGTCACTGTCAAAACCGCTTTTAC 244

QY 2744 TACCTCTAACCTGCAGAGAGCATGACGCGCTTTCTTTTGAATGACTTTGGGATTT 2803  
 Db 245 TACCTCTAACCTGCAGAGAGCATGACGCGCTTTCTTTTGAATGACTTTGGGATTT 304

QY 2804 TTTTAAGCTTTTATTACTTTTCTTTTAACTGTTTCTCTGATGAAACTTTGGGAAG 2863  
 Db 305 TTTTAAGCTTTTATTACTTTTCTTTTAACTGTTTCTCTGATGAAACTTTGGGAAG 363

QY 2864 GGATTAGGAGATCTAGCATTTTATTTCTAGCATGCTTAATTCACGGCTTCCCTATTTTAT 2923  
 Db 364 GGATTAGGAGATCTAGCATTTTATTTCTAGCATGCTTAATTCACGGCTTCCCTATTTTAT 423

QY 2924 ATGTAAATAATTAAGATTTTATTTTATTTTCTTTCTCATAGATATTTTGTAGC-A 2982  
 Db 424 ATGTAAATAATTAAGATTTTATTTTATTTTCTTTCTCATAGATATTTTGTAGCAA 483

QY 2983 TTTTGTGTATTTTGAAGAAATGTGGATAAGTACTTGGTAGTATAAACAGACTCTC 3042  
 Db 484 TTTTGTGTATTTTGAAGAAATGTGGATAAGTACTTGGTAGTATAAACAGACTCTC 543

QY 3043 TGAGATATTGAAATGTTTGGAGATTTTACTTAAACGTACTTTTCAGGAGTGAGC-AAG 3101  
 Db 544 TGAGATATTGAAATGTTTGGAGATTTTACTTAAACGTACTTTTCAGGAGTGAGCAAAG 603

QY 3102 TCCTACTTA-TAAACCTTATTAACCTTT---ATTTTTGGATACCTGTTT 3148  
 Db 604 TCCTACTTA-TAAACCTTATTAACCTTTTAACTTTTGAATACCTGTTT 654

RESULT 11  
 BQ666315  
 LOCUS  
 DEFINITION BQ666315 874 bp mRNA linear EST 21-DEC-2000  
 602119233F1 NIH\_MGC\_56 Homo sapiens cDNA clone IMAGE:4276136 5',  
 mRNA sequence.  
 ACCESSION BQ666315  
 VERSION BQ666315.1 GI:11940210  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 874)  
 NIH-MGC http://imgc.nhl.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.









us-09-873-737a-5.rst

Fri Jan 10 09:31:38 2003

```

QY 1916 CAGTGTGTGGTGGCCGAAACCTTAGGCAACAGCAAACTGTCTATGGCCATTGCTACAAAG 1975
Db 362 CAGTGTGTGTGGTGGCCGAAACCTTAGGCAACAGCAAACTGTCTATGGCCATTGCGACAAG 421
QY 1976 ATTGCCCTACAGATGAACCTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTG 2035
Db 422 ATCGCCCTGCAGATGNACTGCAAGATGGGAGGAGAGCTTTGGAGGGTTGATATGCTCTA 481
QY 2036 AAGCTCGTCATGATCGTTGGCATTCGATTGTTTACCATGACATGACAGCTGGGCGGAGGTCA 2095
Db 482 AAGCTAGCGATGATAGTTGGCATTTGATTCTTATCATGATACCACAGCTGGAGGGAGGTCA 541
QY 2096 ATCGCAGGATTTGTTGCCAGCATCAATGAAGGATG 2131
Db 542 ATTGCAGGATTTGTTGCAAGCATCAATGAAGGATG 577

```

Search completed: January 9, 2003, 21:34:36  
Job time : 3089 secs